KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Modeling and Compensation of Non-Ideal Effects in Microwave Circuits and Systems

Saeed Farsi

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering

July 2014

Modeling and Compensation of Non-Ideal Effects in Microwave Circuits and Systems

Saeed FARSI

Examination committee:
Prof. dr. ir. H. Van Brussel, chair
Prof. dr. ir. D. Schreurs, supervisor
Prof. dr. ir. B. Nauwelaers, co-supervisor
Prof. dr. ir. J. Vandewalle
Prof. dr. ir. M. Steyaert
Prof. dr. ir. S. Pollin
Prof. dr. ir. T. Dhaene (Ghent University)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering

July 2014

ISBN 978-94-6018-850-3

D/2014/7515/74

Acknowledgment

I would first like to express my deepest gratitude to my supervisor, Prof. Dominique Schreurs who supported and guided me through my PhD research. She was always available when I needed her help or feedback. She gave me the opportunity to travel and collaborate with different research groups, which enriched my knowledge and experiences. My warmest thanks go to Prof. Bart Nauwelaers, for his continuous support and encouragement. I am so pleased to have such a gentleman as my co-supervisor.

Thanks to Prof. Vandewalle who provided me helpful articles, comments and discussions during my PhD research. Also thanks to my PhD examination committee, Prof. Steyaert, Prof. Dhaene and Prof. Pollin, for their comments and feedback.

I am grateful for the financial support of FWO Flanders, KU Leuven under the GOA Project, and the Hercules Foundation.

I am deeply indebted to Prof. Peter Asbeck who gave me the opportunity to have an excellent research visit to his wonderful group, HSDG, at the University of California, San Diego (UCSD). To me, he is a nice example of how a man can be extremely knowledgeable, kind and modest at the same time.

I would also like to thank Dr. Kate Remley, for her collaboration through the interesting project that she conducted. Her comprehensive feedback is highly appreciated.

Special thanks to Dr. John Dooley for his collaboration and hospitality in hosting me at the National University of Ireland, Maynooth.

I am grateful to Dr. Prasad Gudem and Dr. Hayg-Taniel Dabag for their constructive discussions, comments and technical information through our meetings at UCSD.

I also wish to express my sincere appreciation to my dear friend Hamed for his help, support and collaboration through this research.

During my stay in Leuven, I have been receiving appreciated help and support from my Iranian friends, particularly Amin, Hadi, Vahid and Majid. I am grateful to them and wish them all the best.

Many thanks to my helpful and nice friends in TELEMIC, Jack, Gustavo and Maciej. Thanks to my roommates Zhongkun, Zhanna, Gian Piero, Raul and Tom, whose company and help are highly appreciated. I am glad to have nice friends and colleagues in TELEMIC from all over the world. Thanks to all of you and I wish you the best of success. I would also like to thank Gabriel and Gian Piero, with whom I had fruitful and pleasant collaborations through supervising their Master's theses.

I sincerely thank Ann Deforce who is always helpful and patient with me and the Telemicers.

My endless gratitude goes to my parents to whom I owe my life.

And finally, I heartfully thank my wife Azam for her kindness, patience, and mostly her love. She has had the most contribution to my PhD research with her invaluable support.

Saeed Farsi
Heverlee, July 2014

Abstract

In top-down design of RF systems starting from the system level, some assumptions and/or simplifications are necessarily made. Descending from the system-level design down into the design of building blocks and components of the system, these assumptions and simplifications are no longer valid. This implies that ideal conditions should be replaced with non-ideal ones. This replacement has some significant consequences, and adversely affects the system performance. Therefore, for a successful realistic design, the non-ideal effects should be characterized and compensated properly. RF systems are usually very sensitive to non-ideal effects. With the advent of next-generation RF systems, particularly telecommunication systems, the need for characterization and compensation of non-ideal effects is getting a huge importance, since this can drastically enhance RF systems' performance. These challenges are the subject of this PhD thesis: non-ideal effects in both transmit band and receive band are investigated. The investigations are conducted in three areas: characterization, modeling and mitigation.

A special case of multisine excitation is proposed to successfully distinguish and characterize non-ideal effects of memory and nonlinearity in and around the transmit band. This ability is achieved by offsetting the tones of a conventional multisine excitation in a specific way. Additionally, another variation of the offset multisine excitation is proposed for identification of nonlinear orders contributing to receive-band non-idealities. Some more techniques are also presented to characterize the deterministic versus stochastic behavior of non-idealities in the receive band.

In the area of modeling, some approaches to model non-ideal systems in and around the transmit band are presented. These novel approaches include a modification to S-functions to enable intermodulation estimation, and a new Volterra-based model. A modification to Volterra-based models is also proposed

to estimate non-ideal effects in the receive band.

Different investigations are also reported in this thesis in the area of mitigating non-ideal effects. Two mitigation (or compensation) techniques are presented based on the aforementioned S-functions and Volterra-based models in the transmit band. These techniques are based on predistortion. A post-processing dithering technique is proposed to mitigate non-ideal effects of data converters. In this approach the input excitation is modulated with a low-frequency waveform to randomize non-ideal effects, and cancel them later after averaging. Different mitigation schemes for canceling receive-band non-idealities are also proposed in this thesis. Some of them are based on modeling in receive band and some are based on blind cancellation. Some of these techniques are performed in the receiver, while some are done in the transmitter by using injection.

All these investigations show that modeling and compensation at the system level, without need for in-depth knowledge of system internal circuitry, is efficient and can be used to significantly improve the system performance.

Samenvatting

Bij het top-down ontwerp van RF-systemen vertrekkende van het systeemniveau, worden noodzakelijkerwijze een aantal veronderstellingen en/of vereenvoudigingen gemaakt. Bij het afdalen van het systeemontwerp naar het ontwerp van de bouwblokken en componenten van het systeem, zijn deze veronderstellingen en vereenvoudigingen niet meer geldig. Dit impliceert dat ideale voorwaarden moeten vervangen worden door niet-ideale. Dit heeft enkele significante gevolgen, en een nadelige invloed op de prestaties van het systeem. Daarom moeten, voor een succesvol, realistisch ontwerp, de niet-idealiteiten gekarakteriseerd en goed gecompenseerd worden. RF-systemen zijn doorgaans zeer gevoelig aan niet-ideale effecten. Met de volgende generatie RF-systemen in het vooruitzicht, in het bijzonder telecommunicatiesystemen, krijgen karakterisatie en compensatie van niet-idealiteiten een enorm belang, omdat zij kunnen bijdragen tot een drastische verhoging van de systeemprestaties. Deze uitdagingen zijn het onderwerp van deze doctoraatsthesis: niet-ideale effecten in zowel de zend- als de ontvangstband worden onderzocht. Dit onderzoek wordt gevoerd in drie domeinen: karakterisatie, modellering en verbetering.

Een bijzondere vorm van multisinusexcitatie wordt voorgesteld om op succesvolle wijze de niet-ideale effecten van geheugen en niet-lineariteit van elkaar te onderscheiden en te karakteriseren, en dit in en rond de zendband. Dit wordt bereikt door de tonen van een conventionele multisinusexcitatie op een bijzondere manier te verschuiven (offset). Een andere, bijkomende variant van de offset-multisinusexcitatie wordt voorgesteld voor de identificatie van de orde van de niet-lineariteiten die bijdragen tot de niet-idealiteiten in de ontvangstband. Enkele andere technieken worden voorgesteld om het deterministische en het stochastisch gedrag van niet-idealiteiten in de ontvangstband van elkaar te onderscheiden en te karakteriseren.

In het domein van de modellering, worden enkele manieren voorgesteld om niet-idealiteiten in en rond de zendband te modelleren. Deze nieuwe benaderingen omhelzen een aanpassing van de S-functies om de schatting van intermodulatie toe te laten, en een nieuw Volterra-gebaseerd model. Een aanpassing van de Volterra-gebaseerde modellen wordt ook voorgesteld om niet-idealiteiten in de ontvangstband te schatten.

Er worden in deze thesis ook verschillende technieken beschreven voor het milderen van niet-ideale effecten. Twee milderings- (of compensatie)technieken worden vermeld, die gebaseerd zijn op de eerder genoemde S-functies en op de Volterra-gebaseerde modellen, beide voor de zendband. Deze technieken zijn gebaseerd op predistorsie. Een post-processing 'dithering'-techniek wordt voorgesteld om niet-ideale effecten te milderen bij gegevensomzetters. In deze benadering wordt de ingangsexcitatie gemoduleerd met een laagfrequente golfvorm om de niet-ideale effecten een willekeurige verdeling te geven, en ze later te verwijderen door uitmiddeling. In deze thesis worden ook verschillende milderingsschema's voorgesteld voor het verwijderen van niet-idealiteiten in de ontvangstband. Sommige van deze technieken zijn gebaseerd op de modellen voor de ontvangstband en sommige zijn gebaseerd op 'blind cancellation'. Sommige van deze technieken worden gerealiseerd in de ontvanger, terwijl anderen worden gerealiseerd in de zender door signaalinjectie.

Al deze onderzoeken tonen aan dat modellering en compensatie op het systeemniveau zonder een gedetailleerde kennis van de interne circuits van het systeem, efficiënt zijn en dat ze kunnen gebruikt worden voor een significante verbetering van de systeemprestaties.

Contents

List of Figures

List of Tables

List of Acronyms

4G	fourth generation
5G	fifth generation
AAA	active antenna array
ACLR	adjacent-channel leakage ratio
ACPR	adjacent-channel power ratio
ADC	analog-to-digital conversion
ADS	advanced design system
AM/AM	amplitude-modulation to amplitude-modulation
AM/PM	amplitude-modulation to phase-modulation
Aux. Rx	auxiliary receiver
AWG	arbitrary waveform generator
BW	bandwidth
BPSK	binary phase-shift keying
CCDF	complementary cumulative distribution function
CDAM	common-differential array model
CDF	cumulative distribution function
CM	core model
CMP	cross-memory polynomial
CW	continuous wave
DAC	digital-to-analog conversion
DDR	dynamic deviation reduction
DDS	direct digital synthesis
DM	differential model
DNL	dynamic nonlinearity
DPD	digital predistortion
DSP	digital signal processing
DSR	disturbance-to-signal ratio
DUT	device under test
EMP	envelope memory polynomial

EVM	error vector magnitude
FDD	frequency-defined device
FFC	feed-forward cancellation
FIFO	first-in first-out
FLOPS	floating-point operation per second
FOM	figure of merit
FPGA	filed-programmable gate array
GCD	greatest common divisor
GMP	generalized memory polynomial
I	in-phase
IF	intermediate frequency
IIP3	input third-order intercept point
IIR	infinite impulse response
IM	intermodulation
IM2	second-order intermodulation
IM3	third-order intermodulation
IMD	intermodulation distortion
IMD3	third-order intermodulation distortion
INL	integral nonlinearity
IQ	in-phase and quadrature
LO	local oscillator
LS	least square
LSB	least significant bit
LTE	long-term evolution
ML	memoryless
MP	memory polynomial
MSE	mean square error
NRMSE	normalized root mean square error
PA	power amplifier
PAPR	peak-to-average power ratio
PH	parallel Hammerstein
PW-DDR	piecewise dynamic deviation reduction
Q	quadrature
QPSK	quadrature phase-shift keying
RI plot	real-imaginary plot
RLS	recursive least square
Rx	receive
RxBN	receive-band noise
SAW	surface acoustic wave
SDR	software-defined radio
SFDR	spurious-free dynamic range

SQNR	signal-to-quantization-noise ratio
SVD	singular-value decomposition
TWT	traveling-wave tube
Tx	transmit
TxSI	transmitter signal injection
UART	universal asynchronous receiver/transmitter
VSA	vector signal analyzer
VSG	vector signal generator
WCDMA	wideband code division multiple access

Chapter 1

Introduction

1.1 Introduction

RF systems are everywhere in our daily life: mobile telecommunications, Wi-Fi and Bluetooth data connections and TV/radio broadcast are just a few examples of the RF system applications. Many assumptions and simplification are made in the design of such systems; the systems are assumed to be ideal in many senses. For example, sometimes they are assumed to be *linear* or *memoryless*. However, these assumptions are usually violated in reality, and there are often considerable deviations from such assumptions. If the designer or the system engineer is not aware of these non-idealities, the designed system may simply not work desirably. To deal with the non-ideality problem, they should be defined and quantified well.

This chapter represents the framework of this PhD research which was conducted for identifying and dealing with such non-idealities. The context in which this PhD study was conducted is explained in Section 1.2. Section 1.3 introduces the most important non-idealities observed in RF systems. The next section, Section 1.4 discusses the state-of-the-art trends and approaches to deal with and mitigate the non-idealities. Since the focus of this PhD thesis was at the system level, only the system-level approaches are covered. The targets of this PhD research are presented in Section 1.5, and finally, the thesis organization is explained in Section 1.6.

1.2 Motivation and applications of the work

As explained in the previous section, non-idealities can be present and degrade performance in various RF systems. Although the discussions made in this thesis are general enough to be applied in different applications, the special interest of this PhD study was telecommunication systems. Exponentially increasing demand for higher data rates caused by rising number of clients and increasing requested data rate per client requires continuous enhancement in telecommunication services, which in turn lead to designing higher-quality and more complicated telecommunication standards. For example, telecommunication operators are switching to modern-generation standards and technologies, such as the emerging fourth generation (4G) and potentially the future fifth generation (5G) telecommunications.

Previously, the design of the RF front-ends of telecommunication systems were restricted to analog design. Different architectures and techniques were proposed to improve the performance of RF front-ends in terms of power efficiency, linearity, and low-noise operation. However, the new generations of telecommunication systems require more stringent specifications that were not achievable before by using analog design only. With breathtaking development in digital circuits which has led to compact low-power high-speed digital architecture, it is now possible to incorporate digital blocks into RF front-ends. With such development, a mixture of digital and analog schemes can be deployed to improve the front-end performance. Digital schemes can outperform analog schemes in terms of accuracy, reliability, flexibility and reconfigurability, and power efficiency. Therefore, as the digital circuits are getting higher performance, their applications in the RF front-end markets are increasing and spreading. Their combination with analog schemes can yield outstanding results. For example, they can be combined with analog schemes that have high power efficiency but poor linearity, in order to compensate their nonlinearity. The result is a circuit with high efficiency and linearity at the same time. Such combined schemes can easily be implemented in base stations. In handsets, however, there are still challenges to pave the way for efficient and complete implementation of those digital schemes. Simplified digital schemes are already being used in handsets. Nevertheless, efforts are made to improve the digital schemes in both software and hardware. Clearly, improvement in hardware paves the way for implementing more complicated algorithms (software).

One of the ultimate targets in the improvement of digital circuits is to achieve a wideband software-defined radio (SDR) system, in which the signal is treated in digital domain in almost all of the system blocks, except for the very front end which could be a minimal set of filters and amplifiers only [9, 10]. This

architecture provides excellent flexibility for the telecommunication system. In such an SDR system, the superiority of digital schemes for performance improvement is most felt.

This thesis focuses on algorithms and schemes that can be used to characterize, model and digitally improve the RF front-end performance. More specifically, the thesis deals with non-ideal effects which are introduced in the following and will be comprehensively dealt with in the next chapters. Note that, as already stated, the proposed approaches are general enough to be applied to RF applications other than telecommunications.

1.3 Non-ideal effects in RF systems

Any phenomenon that violates the ideality assumption of RF systems can degrade their performance. We refer to these phenomena as *non-ideal effects*. In this section three important types of non-ideal effects that usually occur in RF systems will be explained: nonlinearity, memory effects and in-phase and quadrature (IQ) imbalance.

1.3.1 Nonlinearity

Definition

A linear system is the one that has a linear relationship between its input $x(t)$ and output $y(t)$ that can be illustrated as a function $f\{.\}$:

$$\begin{aligned} y(t) &= f\{x(t-\tau)\} \\ f\{a_1x_1(t)+a_2x_2(t)\} &= a_1f\{x_1(t)\}+a_2f\{x_2(t)\}, \end{aligned} \tag{1.1}$$

where τ is a constant output delay with respect to the input, and a_1 and a_2 are constant coefficients. It can be shown that in a time-invariant system, a linear function has the form:

$$f\{x(t)\} = ax(t)+b, \tag{1.2}$$

in which a and b are constant coefficients. In RF systems, the coefficient a is usually referred to as *gain*. The coefficient b determines the bias point, or operating point, around which the output is varying. This coefficient usually adds only a constant value to the output and does not bear any variation of the

response. Hence, b is usually disregarded in the analysis of RF systems, since it is usually assumed that the operating point is fixed. The simplified relationship in discrete-time is then:

$$y_n = ax_n, \tag{1.3}$$

ignoring the constant delay difference between the input and output. If the delay difference n_τ is to be considered, x_n will be replaced with x_{n-n_τ}. The signals y_n and x_n denote the output and input values at the sample time n, $y(n)$ and $x(n)$ respectively. Note that in baseband representation, the gain a can be complex-valued.

Contrary to the linear system, the nonlinear system has a non-constant gain g(t) which depends on the amplitude of the input:

$$y_n = g(|x_n|)x_n, \tag{1.4}$$

Figure 1.1 shows the typical plots for the output power versus the input power in linear and nonlinear systems.

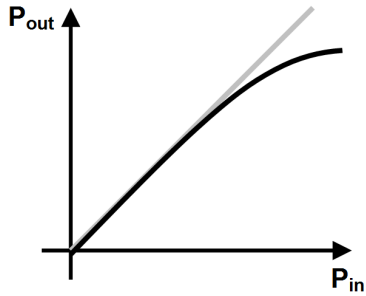


Figure 1.1: Typical output power versus input power in linear (gray) and nonlinear (black) systems.

Weak vs. hard nonlinearity

Generally, there are two types of nonlinearity: *hard nonlinearity* and *mild nonlinearity* [11, 12]. The former includes hard clipping of the waveform due to some mechanism such as cut-off, breakdown or saturation. However, the latter, also known as *weak nonlinearity*, includes weak nonlinearities that exist even with small signal variation around the operating point. Figure 1.2 shows these two types of nonlinearity in a typical characteristic curve of an amplifier,

disregarding memory effects. As can be seen in the figure, the characteristic curve is not straight, even in its middle. This weak deviation from being a straight line (dotted line) is the mild nonlinearity, whereas the clippings at the two ends of the curve are hard nonlinearities. The behavioral models discussed in this chapter are used to model mild and close-to-hard nonlinearities, and usually fail to have acceptable performance for very hard nonlinearities [12, 13].

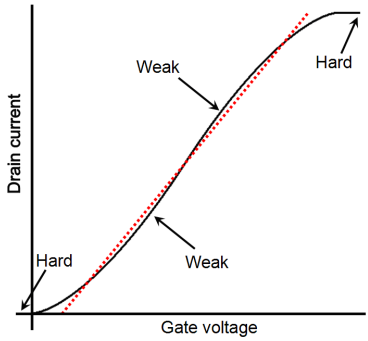


Figure 1.2: Demonstration of weak and hard nonlinearities in a characteristic plot of a field-effect transistor.

Harmonic generation and intermodulation

Among the most important consequences of nonlinearity are harmonic generation and intermodulation. Consider a tone at the frequency f_c, mathematically expressed by $e^{j2\pi f_c t}$. When a nonlinear block of the form $a_1x+a_2x^2+a_3x^3+\ldots$ is excited by such a tone, the output will be $a_1e^{j2\pi(1f_c)t}+a_2e^{j2\pi(2f_c)t}+a_3e^{j2\pi(3f_c)t}+\ldots$. In other words, the harmonics of the input frequency f_c, namely $2f_c$ and $3f_c$ and $\ldots$, are generated.

Another phenomenon is *intermodulation*. This phenomenon happens when two frequency components of the input signal mix together due to nonlinearity. This mixing product can lie around the carrier frequency, DC or higher harmonics. Those products around DC and around the carrier frequency are of special ineterst, as will be seen later in this PhD thesis. Figure 1.3 illustrates the harmonic and intermodulation products of a nonlinear block with linear, second-order and third-order terms, excited by a three-tone excitation. The figure shows that each harmonic or intermodulation product comes from which nonlinearity

order. The generation of intermodulation products around the carrier frequency is usually referred to as *spectral regrowth*. In Fig. 1.3 spectral regrowth is caused by third-order nonlinearity only. Spectral regrowth are generally created by odd-order nonlinearities, since even-order mixing products do not lie around the carrier frequency.

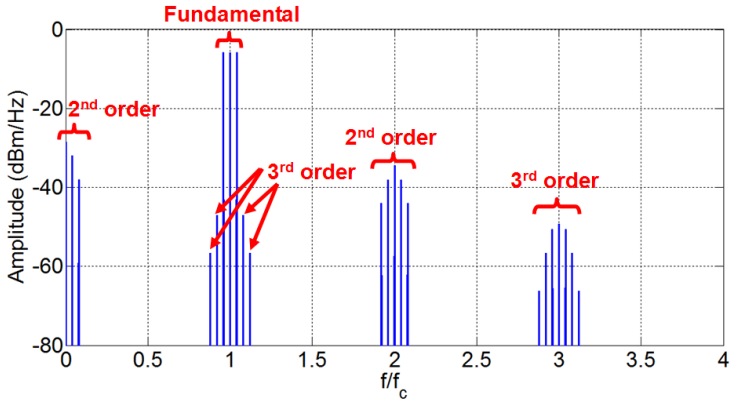


Figure 1.3: Generation of harmonics and intermodulation products due to nonlinearity excited by a three-tone signal.

A special case of intermodulation is of special interest, when two separate bands are simultaneously present at the input of a nonlinear block. In this case, not only the above-mentioned frequency components of each band are intermodulated together, but the frequency components of the different bands are intermodulated also. This type of mixing is called *cross-modulation*, and is shown in Fig. 1.4. In this figure, intermodulation products of each band along with the cross-modulation products can be seen. It is clear that the frequency location of cross-modulation products can be different from that of intermodulation products, as is the case in Fig. 1.4.

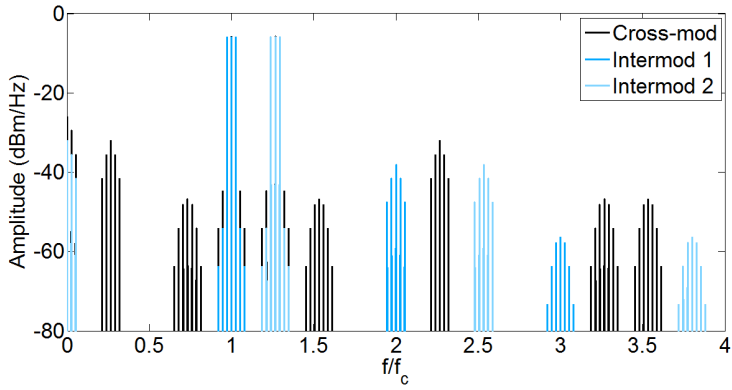


Figure 1.4: Generation of intermodulation and cross-modulation products when two separate bands are simultaneously present at nonlinearity input.

1.3.2 Memory effects

Definition

In general, the input-output relationship of a system could be *static* or *dynamic*. Here, the static or *memoryless* relationship is the one that has the form:

$$y_n = f(|x_n|). \tag{1.5}$$

In other words, in the static case, the output is function only of the current input. In the opposite, the dynamic or memory-including relationship is the one that relates the output to not only the current input, but also to the previous values of the input, i.e.,

$$y_n = f(x_n, x_{n-1}, x_{n-2}, \ldots). \tag{1.6}$$

Under this assumption, the system is said to have *memory*. Memory in the frequency domain is translated to frequency dependency or *dispersion*. The simplest type of memory is linear memory or *linear distortion*. The output of a system with linear memory has the form:

$$y_n = h * x_n = \sum_m h_m \cdot x_{n-m} \tag{1.7}$$

in the time domain. The output is the convolution of the impulse response with the input $x(.)$. In the frequency domain, it is the multiplication of the Fourier

transformations of $h(.)$ and $x(.)$, $H(\omega)$ and $X(\omega)$, given by:

$$Y(\omega) = H(\omega) \cdot X(\omega). \tag{1.8}$$

The abovementioned distortion is linear, because the coefficients of x_{n-m} in (1.7) do not depend on the input value. As opposed to this type of memory, there is *nonlinear memory*. In such a case, the coefficients are dependent on the input values. For example:

$$\begin{aligned} y_n &= \sum_m g_m(x_{n-m}) \cdot x_{n-m} \\ g_m(x_n) &= \sum_k |x_n|^k, \end{aligned} \tag{1.9}$$

in which the coefficients $g_m(\cdot)$ are themselves functions of the input, has nonlinear memory.

There is an intermediate condition in between the memoryless and memory-including cases, called the *quasistatic* or *quasi-memoryless* case [14]. In such a condition, both the output amplitude and output phase are assumed to be a static nonlinear function of the input amplitude, given by [14, 15]:

$$|y_n| = f(|x_n|) \tag{1.10}$$

$$\angle y_n = q(|x_n|). \tag{1.11}$$

Note that the difference between the quasistatic and purely static condition is that, in the latter case the phase is also dependent on the input amplitude, whereas in the former case, only the gain is dependent. In other words, in the static case, we have only amplitude modulation to amplitude modulation (AM/AM) conversion, or only eq. (1.10).

Equation (1.11) is called amplitude modulation to phase modulation (AM/PM). More explanation will be provided in Section 2.3.2.

Long-term vs. short-term memory

Memory effects may have short or long time constants, called short-term or long-term memory effects, respectively [16, 17]. Short-term memory effects are usually caused by frequency dependence of device internal parasitics, matching networks and/or mismatching between different parts in the RF path. They are

often in the scale of nanoseconds, and can be usually diagnosed as the amplitude variation of the main-band output spectrum. Long-term memory effects are often in the scale of milliseconds or microseconds, and are known to have three major sources: non-ideal bias circuit [18, 19], thermal effects [18, 19, 20, 21] and trapping effects [21, 22].

Long-term memory effects in a power amplifier (PA) can be described by a feedback mechanism [23], as shown in Figure 1.5. In this mechanism, an envelope signal is created by even-order nonlinearities of $F_{NL}(\omega)$, which represents the static nonlinearity of the device. This low-pass signal then manipulates either the thermal or the electrical DC operating point. The manipulation mixes the generated envelope signal with the existing signal, resulting in some non-ideal effects around the fundamental frequency and its harmonics (Long-term memory effects usually are referred to the effects around the fundamental frequency, not the harmonics). Since the thermal or electrical manipulation is frequency-dependent, reflected in $F_{FB}(\omega)$, it leads to memory effects. Considering that the thermal and DC operating point variation have large time constants compared to the RF signal, these memory effects are long-term. The two input and output blocks, $H(\omega)$ and $O(\omega)$, are representatives for linear distortion effects at the input and output of the PA circuit.

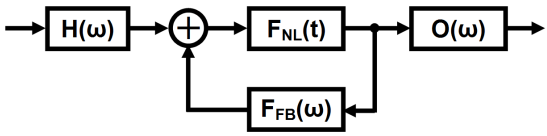


Figure 1.5: Feedback mechanism describing long-term memory effects in a PA circuit.

As mentioned above, along with non-ideal biasd circuits and thermal effects, trapping effects can also cause long-term memory. As was mentioned in [17]:

> At low frequencies, electrons or holes can be trapped in surface states or in the bulk material due to imperfections in the semiconductor structure. As the time constants of the trapping and releasing processes are relatively large, trapping mainly affects low-frequency signals. Holes can be generated by impact ionisation in the channel and then become trapped. Hole-trapping time constants are therefore often observed to be a strong function of the drain-source bias, in accordance with impact-ionisation characteristics.

1.3.3 IQ imbalance

Many RF systems use IQ modulator/demodulator for up- or down-conversion [24]. In these demodulators/demodulators it is assumed that the signals in the in-phase (I) and quadrature (Q) paths are equal in amplitude and have 90 degrees difference in phase. However, these assumptions are not perfectly fulfilled in practice, and there are always deviations from these two assumptions [25, 26, 27]. These deviations are called IQ imbalance. In other words, the I and Q paths can have unequal amplitudes (amplitude imbalance) or a phase difference other than 90 degrees (phase imbalance). IQ imbalance has degrading effects in the system performance and should be compensated in many applications [25, 26, 27].

1.4 State-of-the-art trends in non-ideality mitigation

There have been increasing efforts to mitigate the aforementioned non-ideal effects. In addition to the established techniques available in this field of research, some new techniques are emerging. These new approaches are in response to emerging more stringent requirements. In many applications, such as the available third-generation (3G) and emerging fourth-generation (4G) telecommunications technologies, the requirements with the utmost importance are higher data rates and more efficient power consumption. Higher data rates in turn demand for the data signals with higher peak-to-average power ratio (PAPR) and/or larger bandwidth. Higher PAPR leads to more pronounced nonlinearity. To mitigate this nonlinearity while keeping the system components, in particular power amplifiers, at power efficient operating points, compensation techniques have been proposed [28]. For emerging standards, these techniques require to have higher performance, and their implementations should be power-efficient. Power-efficient implementation requires simple techniques (in terms of architecture) in order to have simple hardware for lower power consumption [29, 30]. Hence, the trend is to find simplified techniques for higher power efficiency, while keeping them high-performance.

Higher data rates also demand for larger bandwidth, or multiband schemes. In systems with large-bandwidth signals, some more considerations should be given for compensation. For example, the sampling speed limitation could be a problem that should be dealt with [31, 32, 33]. For example, memory effects are more pronounced, and the emerging compensation techniques employed in such systems should be able to estimate large memory effects accurately. In systems

with multiband signals, multiband compensation schemes are attracting more and more attention. These techniques should consider not only the non-ideal effects of each band in and around itself (intermodulation), but also the non-ideal interaction effects between bands that may show up in different bands (cross-modulation) [34, 35, 36].

Another class of emerging compensation techniques are involved in power-efficient RF circuits. More specifically, power-efficient amplifier designs such as Doherty [37], envelop tracking (ET) [38], and power tracking (PT) power amplifiers [39] have been increasingly used in recent RF systems. These amplifiers require their own specific compensation techniques [38, 40, 41].

One of the most recent research area in non-ideal compensation is on RF transceivers, where due to non-ideal effects the transmit signal degrades the receiver performance [42, 43, 44, 45, 46]. Such degradation is more pronounced in recent and future telecommunication standards, where the transmit-receive frequency difference is closer to the signal bandwidth. Furthermore, in multiband applications, cross-modulation products of different bands may lie in the receive band [47]. Another case of interest in this research can be the one with ET or PT amplifiers, which can lead to more pronounced transmit signal "emission" to the receive-band. The receive-band emission compensation techniques will get much high importance in future telecommunications.

Related to this research area, some waveform "engineering" can also be performed [48, 49, 50]. The aim of this waveform tailoring is to better characterize the system non-ideal effects. One of the examples is multisine design, i.e., setting the amplitudes and phases of the tones to obtain a multisine excitation that can excite the desired behavior of the system [49, 50].

1.5 Research aims of thesis

One of the key tools for non-ideality compensation and performance improvement is signal processing. Signal processing can be employed in RF systems to compensate nonlinearity and memory effects, as well as some other non-ideal effects, to a high degree. The main subject of this PhD thesis is about approaches that use signal processing for system performance improvement. In order to do this enhancement at the system level, we need to have system-level models of the system behavior. More specifically, we require system-level model of non-ideal effects in order to be able to compensate them with signal processing.

This PhD thesis presents different techniques of system-level modeling of non-ideal effects. Some compensation methods are also proposed based on these techniques to improve the system performance. The proposed compensation techniques are in two categories: techniques for non-ideality modeling and mitigation around the transmit band, and those for non-ideality mitigation in the receive band.

As mentioned in the previous section, waveform engineering for excitation signals is beneficial to be able to diagnose non-ideal effects. This PhD thesis also aims at waveform engineering for better characterization of non-ideal effects, which can contribute to better modeling and compensation. All the proposed methods and techniques are tried to be validated by simulations and/or experiments.

1.6 Outline of thesis

Figure 1.6 summarizes the research activities conducted through the PhD study. As can be seen, to tackle the non-ideal effects problem investigations are done in three areas: characterization, modeling, and mitigation or compensation. This PhD thesis covers all the three areas. The research activities can also be categorized based on whether they consider spectrum around the transmit band or the receive band, as shown in Fig. 1.6.

Chapter 2 explains some concepts and definitions used in system-level modeling. Furthermore, some widely used models which are the basis of the modeling approaches presented in this PhD thesis are explained in that chapter. Then, based on the introduced models a technique which is the framework for compensation of non-ideal effects in this PhD thesis is discussed.

Different techniques for compensation of non-ideal effects are proposed in Chapter 3. The first technique employes a modified version of S-function [1] which are explained in Chapter 2. A novel system-level model is also presented in that chapter. Then, based on the proposed model, a compensation technique is explained, which employes a setup developed in the TELEMIC group, KU Leuven. As another application, an array of amplifiers used in an active antenna array are modeled. This model can be used for the compensation of non-ideal effects in the array.

As mentioned before, for successful modeling and compensation of non-ideal effects, a good characterization of such effects is highly helpful. Chapter 4 proposes the use of a novel excitation, called offset multisine, to have enhanced

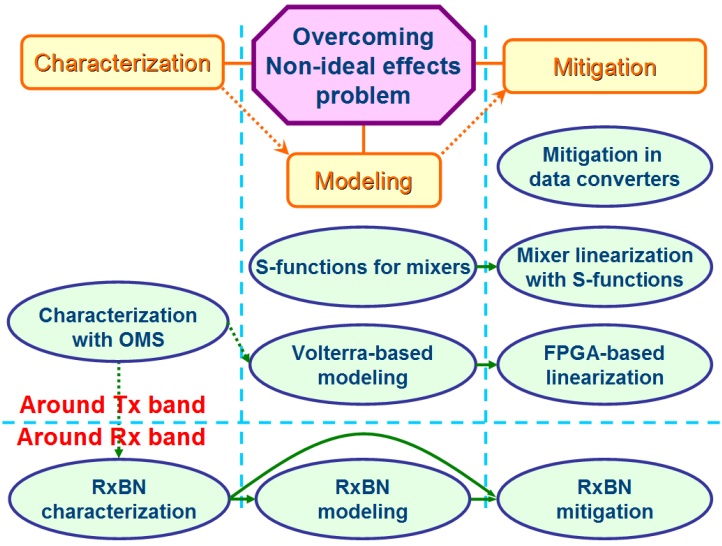


Figure 1.6: Outline and interconnections of research activities covered by this PhD thesis.

characterization and identification of non-ideal effects. The approach is tested for different applications.

Apart from the non-ideal effects discussed in Chapter 2, there are some other non-ideal effects specific to analog-to-digital and digital-to-analog data converters. Corresponding to these effects, some techniques for their compensation are explained in Chapter 5. A novel approach for such compensation is also presented and validated with some experiments in that chapter. The approach is based on waveform engineering and post-processing.

As explained in Section 1.4, non-ideal effects in RF transceivers cause distortion not only in and around the transmit band, but also in the receive band. This distortion causes desensitization of the receiver, as explained in Chapter 6. In that chapter, the mechanisms that cause this type of distortion are discussed, and some methods are proposed to quantify and characterize these mechanisms. A modeling approach is also proposed for modeling of this receive-band distortion. Furthermore, some techniques are presented for mitigation of this distortion. Finally, in Chapter 7, some conclusions are drawn based on what is presented in this PhD thesis, the contributions of the thesis are listed, and some directions for future research are presented.

Chapter 2

Behavioral Modeling and Compensation

2.1 Introduction

Prior to presenting the contributions that were made during the PhD study of the author, some definition and introductions are required. The materials of this chapter is the basis of the next chapters of the PhD thesis. This chapter introduces some approaches that are used at the system level to model non-ideal effects, with a special focus on one of their most widely used types, namely *Vollterra series*. Some important figures of merit to assess the performance of the modelds are introduced, and a technique to reduce the non-ideal effects, called digital predistortion (DPD), is explained. This technique is based on the models explained in this chapter, more specifically the Volterra series.

In this chapter and also the entire PhD thesis, the discussions are all in the discrete-time domain, unless otherwise explicitly stated. However, the continuous form of the formulations and relationships can be derived easily by corresponding conversions, for example, conversion of summation to integration.

Section 2.2 introduces the baseband representation which is used in this PhD thesis. Section 2.3 discussed some frequently used system-level models that are used to model non-ideal effects. Some figures of merits used for evaluating those behavioral models are introduced in Section 2.4. Based on the models

introduced in this chapter some techniques to compensate non-ideal effects are presented in Section 2.5. Finally, conclusions are drawn in Section 2.6.

2.2 Baseband representation

RF signals used in practice are band-limited. Limited bandwidth leads to the definition of a type of signal representation, called *baseband*, which makes analysis simpler. In fact, by using baseband representation, we skip considering the RF carrier in analysis which does not contain any information, but is used only for carrying information. The baseband representation is widely used in system-level simulation and analysis, and for modeling and linearization purposes as will be shown in this PhD thesis.

If a signal $x(t)$ is band-limited around the center (or carrier) RF frequency f_c, it can be described by:

$$x(t) = \mathrm{Re}\{x_{LP}(t)e^{2\pi f_c t}\}, \tag{2.1}$$

where $x_{LP}(t)$ is a complex lowpass signal around DC with the bandwidth the same as that of $x(t)$. The lowpass $x_{LP}(t)$ is the baseband form of $x(t)$, and has a spectrum shape ($X_{LP}(\omega)$) which is the same as that of $x(t)$ ($X(\omega)$), as shown in Figure 2.1. As can be seen in the figure, $X_{LP}(\omega)$ is asymmetric, which means that $x_{LP}(t)$ is complex-valued. For brevity in this PhD thesis, $x_{LP}(t)$ is denoted by $x(t)$ itself unless otherwise explicitly stated.

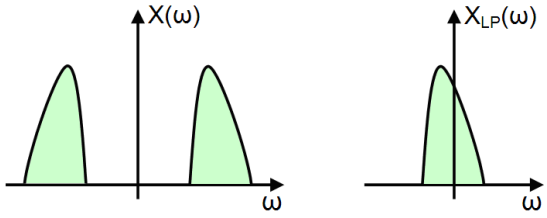


Figure 2.1: Spectra relationship between $X(\omega)$ and $X_{LP}(\omega)$.

2.3 Behavioral modeling

Modeling and simulation is a key part of designing a circuit/system. Without the knowledge of how the circuit would respond to an excitation, the design cannot be done. There are various approaches to modeling. Starting from low level to high level, the models can be physics-based, equivalent-circuit or system level [51]. Physics-based models try to predict the system response based on the physical mechanisms that act in the system. Therefore, they require detailed physical information about the system. Such approaches are very computation-intensive. Going to a higher hierarchical level, equivalent-circuit models try to find an equivalent circuit of the system for prediction purposes. Thus, they require considering different equivalent elements for different parts and different mechanisms of the system. As compared to the physics-based models, they require fewer details, hence are more computationally efficient. At the highest hierarchical level, there are system-level models. Such models try to mathematically formulate (or predict) the system response without going into physical details. In other words, the system-level models disregard the internal mechanisms of the system that lead to the corresponding output. They consider only the output itself. For such prediction, the system-level models make a mathematical relationship between the input and the "observed" output. Since the models are based only on such observation, or system *behavior*, they are called *behavioral models*.

As mentioned earlier, behavioral models do not require knowledge about the internal circuitry, and they are based only on input-output observations. Therefore they are highly efficient in terms of computation time and complexity. Their accuracy is sufficient in many system-level applications. Because of their compact form and sufficient accuracy they are the first choice for system-level analysis and compensation of non-ideal effects, as will be discussed in Section 2.5 and the next chapters. Furthermore, since they are built based on input-output observations, they can be highly accurate for the excitations of the same type. However, due to the same reason, they are potentially signal-specific. Their acceptable performance should be validated with unused dataset yet similar to the one used for identification. Behavioral models belong to various classes, some of which are described in the following subsections.

2.3.1 S-functions

S-functions are generalization of S-parameters [52]. They characterize nonlinear circuits in terms of their response at fundamental and harmonic frequencies,

whereas S-parameters are used to characterize linear circuits in the frequency domain.

S-functions model a device that is operating in a nonlinear steady state which is determined by a large single-tone input to the nonlinear circuit. The other tones available at the input and output at fundamental and harmonic frequencies are assumed to be much smaller than that large tone. These additional small tones can be generated by nonlinearity and reflections. Under these assumptions, a multi-harmonic linearization approximation around the operating point is made: the center value of each tone is determined by the large tone, and the additional generated small tones are perturbations that change the values linearly. The model relates A-waves and B-waves together, where A- and B- waves are the incident and reflected travelling pseudo-waves defined as:

$$\begin{aligned} A &= \frac{V + Z_c I}{2} \\ B &= \frac{V - Z_c I}{2}. \end{aligned} \tag{2.2}$$

where V, I and Z_c are the voltage, current and the characteristic impedance of the port, respectively. The S-functions model is mathematically illustrated by [52]:

$$\begin{aligned} B_{pm}(|A_{11}|, f) = &\sum_q \sum_n Sf_{pmqn}(|A_{11}, f) \cdot A_{qn} \cdot \exp(j(m-n) \cdot \phi(A_{11})) \\ &+ \sum_q \sum_n Sfc_{pmqn}(|A_{11}, f) \cdot A^*_{qn} \cdot \exp(j(m+n) \cdot \phi(A_{11})), \end{aligned} \tag{2.3}$$

where A_{pm} and B_{pm} are the A- and B- waves. Sf_{pmqn} and Sfc_{pmqn} are the S-functions which are the complex coefficients whose values depend on the magnitude and frequency of the large tone. The parameter p denotes the port number at which the reflected wave is to be calculated, and m is the harmonic number of the reflected wave. The parameter q denotes the port number to which the A wave is incident, and n is the harmonic number of the incident wave. Equation (2.3) assumes that the reflected wave is a nonlinear function of A_{11}, but is linear with respect to the other A-waves. This assumption is valid as long as these waves are much smaller than A_{11} so that they do not change the operating point. The equation also shows dependency of reflected waves to not only the incident waves, but also their conjugate values A^*_{qn}, modeling the non-analyticity of the describing function [52]. Another point is the dependency of the reflected waves to the absolute phase value of A_{11}. This is the direct result of time invariance assumption [53].

Properties of S-functions

Because of its form, the S-functions model is well-suited for harmonic balance simulation. The formulation already provides required relationships between the tones at the fundamental and harmonic frequencies. Furthermore, the S-functions model provides high-accuracy estimation for different nonlinear characteristics, including compression, AM/PM, harmonics, load-pull, and time-domain waveforms [52]. The model can accurately predict the reflected B-wave in nonlinear systems with mismatches at both the fundamental and harmonics. This capability enables the accurate total distortion estimation of circuits composed of a cascade of nonlinear components with known S-functions for each of them.

On the other hand, S-functions models are accurate as long as the assumption of the dominant large tone is valid. If one or more of the A-waves other than A_{11} are not much smaller than A_{11}, the assumption is violated and the accuracy is not guaranteed. The violation conditions include large mismatches and modulated excitations. Failure in the case of large mismatches is common with other behavioral models as well. However, failure in the case of modulated excitations is not the case in many other models, and is a considerable drawback. S-functions can be used for modulated excitation conditions under the assumption of quasistatic situation. This assumption holds as long as the time of change in the modulating signal is much larger than the physical time constants that are inherent in the system. However, this is not the situation in many wideband applications, for example when long-term memory effects are not negligible. There are efforts to include long term memory in the S-functions model [54, 55]. Nevertheless, the derived models are too complicated and still not general enough to be used in many applications.

2.3.2 AM/AM and AM/PM models

When the system has a little memory, it can be considered as quasi-static or quasi-memoryless. In such a situation, equations (1.10) and (1.11) hold, which are representatives of AM/AM and AM/PM conversions [14, 15]. These equations state that both the gain and phase relationships between the input and output depend on the input power level, but not on the signal modulation. These relationships are typically illustrated by a plot, as in Figure 2.2. If the system was memoryless (static case), only the AM/AM curve would be defined and the phase would be independent of the input power level, as shown in

Figure 2.2. In other words, the AM/PM curve would be a horizontal straight line.

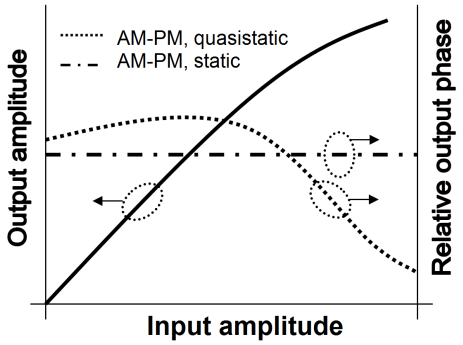


Figure 2.2: Typical AM/AM and AM/PM curves of a quasistatic system.

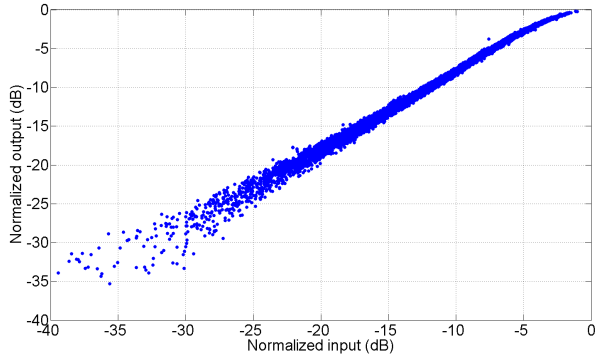
One of the most well-known AM/AM and AM/PM models is the Saleh model. It was originally used for traveling-wave tube (TWT). It relates the gain and phase of the transfer function to the input amplitude, through rational polynomial functions [13, 56].

When the amount of memory increases, the AM/AM and AM/PM models start to fail in providing acceptable estimation. The AM/AM and AM/PM plots are not single curves anymore and cannot be defined by equations like (1.10) and (1.11). In such cases the plots will be in the form of a set of scattered dots, as shown in Figure 2.3.

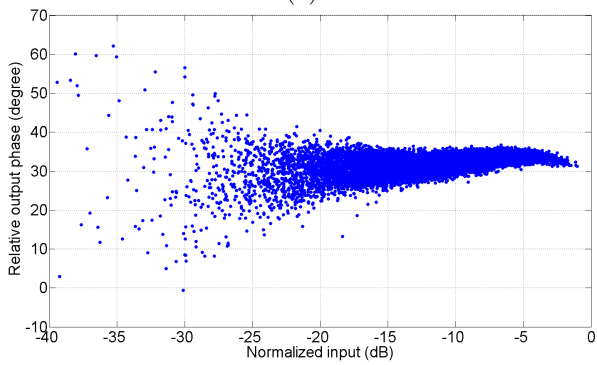
2.3.3 Volterra-based models

Volterra series and its derivations are widely used for behavioral modeling. Its baseband discrete-time form can be represented by [12]:

$$y_n = \sum_{p=1}^{P} \sum_{m_1} \cdots \sum_{m_{2p+1}} \left[h_{2p+1}(m_1, \ldots, m_{2p+1}) \prod_{i=1}^{p+1} x_{n-m_i} \prod_{j=p+2}^{2p+1} x^*_{n-m_j} \right], \quad (2.4)$$



(a)



(b)

Figure 2.3: Typical a) AM/AM and b) AM/PM curves of a system with non-negligible memory.

where $h_k(\cdot)$ is the k^{th}-order Volterra kernel. Take the k^{th}-order term for instance, represented as:

$$y_{nk} = \sum_{m_1} \sum_{m_2} \cdots \sum_{m_k} \left[h_k(m_1, \ldots, m_{2p+1}) x_{n-m_1} \ldots x_{n-m_{p+1}} x^*_{n-m_{p+2}} \ldots x^*_{n-m_{2p+1}} \right]$$

$$k = 2p + 1. \tag{2.5}$$

Considering what is explained in section 1.3.2, this term includes nonlinear memory effects. If $k = 1$, the term is the linear convolution, and thus describes the linear behaviour of the system, including the linear distortion. The baseband representation of Volterra series contains only odd orders, since the even orders do not lead to any frequency product around the fundamental frequency [57].

Also the series is truncated at the order $2P+1$ for practical implementation. It is assumed that the contributions of higher orders are insignificant.

Identification of Volterra-based models

The unknown parameters of eq. (2.4) are the coefficients $h_{2p+1}(m_1,\ldots,m_{2p+1})$. All these coefficients have first order, which leads to a linear equation with respect to the unknown parameters. Therefore, to find the unknown values, a sufficient number (more than the number of unknown values) of eq. (2.4) for different n values should be considered. This system of linear equations (2.4) with varying n can be written as:

$$\mathbf{Y} = \mathbf{X}\mathbf{b}, \tag{2.6}$$

where $\mathbf{Y}$ is the vector of output samples y_n, $\mathbf{b}$ is the vector of unknown coefficients $h_{2p+1}(m_1,\ldots,m_{2p+1})$, and $\mathbf{X}$ is the input matrix consisting of elements $\prod_{i=1}^{p+1} x_{n-m_i} \prod_{j=p+2}^{2p+1} x^*_{n-m_j}$, with different n values. Such a system of linear equations can be solved by using least square (LS) methods [58].

The number of coefficients in eq. (2.4) increases exponentially with increasing the truncation order $K = 2P+1$. In many applications, the orders up to at least five or seven are required to have acceptable estimation of the system nonlinearity. This requirement leads to a huge number of coefficients for computation. Apart from increased complexity, the large number of coefficients may lead to overfitting.

A solution to this complexity problem is to make simplification in Volterra series. There is an enormous number of papers in the literature reporting different variations and simplifications of Volterra series, e.g, [59, 60, 8, 61, 62, 63, 64]. In this subsection, only some of the most widely used and most accurate ones with suitable compromise between accuracy and simplicity will be addressed.

Effect of cross terms

Some of the terms in (2.4) are multiplication products of the same input sample, i.e., $h_k(m,m,\ldots,m)$. The other terms are the multiplication products of different input samples. These terms are termed as *cross terms*. Cross terms have been proven to cause improvement in modeling capability as compared to the Volterra-based models without cross terms [64]. This improvement is due to the fact that cross terms actually imply that different input samples

from different time instances are mixed together in the nonlinear circuit. This mixing may occur due to, for example, feedback or multipath.

Including even-order terms

Even-order terms in baseband have the form

$$h_{2p}(m_1,\ldots,m_{2p+1})\prod_{i=1}^{q}x_{n-m_i}\prod_{j=q+1}^{2p}x^*_{n-m_j}, \tag{2.7}$$

where p and q can be any integer number. Starting from the RF representation of Volterra series, i.e.,

$$y_n=\sum_{k=1}^{K}\sum_{m_1}\cdots\sum_{m_k}\left[h_k(m_1,\ldots,m_k)\prod_{i=1}^{k}x_{n-m_i}\right], \tag{2.8}$$

and keeping only the terms that generate frequency components around the fundamental frequency, equation (2.4) can be derived [57]. Such a derivation does not include even-order terms. In other words, even-order terms cannot be derived from the Cartesian representation of Volterra series. However, it was shown in [19] that starting from the polar representation on Volterra series and with some assumption regarding the nonlinear mechanisms in the circuit, even-order terms can be included in the formulation. The assumed mechanisms in [19] came from long-term memory effects. It has been shown in the literature that including even-order terms in Volterra-based models improves the modeling accuracy in many applications [65, 66].

Wiener and Hammerstein models

For simplification, it may be assumed that the nonlinearity and memory mechanism of the system are separable. In such a case, a model can consist of two cascaded blocks on nonlinearity (N) and linear memory (G) as shown in Figure 2.4. As the figure shows, if N precedes G, the model is a Hammerstein model, and if N follows G, it is a Wiener model. The identification of the Wiener model is more difficult, because the resulting equations are nonlinear with respect to the coefficients to be found, whereas they are linear in Hammerstein modeling.

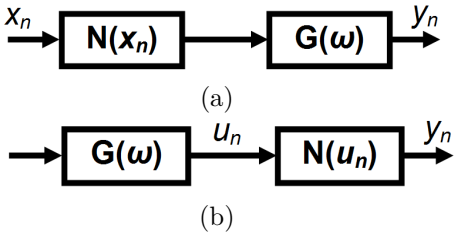


Figure 2.4: (a) Wiener and (b) Hammerstein models

Memory polynomial model

The memory polynomial (MP) or parallel Hammerstein (PH) model is one of the most widely used Volterra-based models because of its combined compactness and high performance. It is represented by [61]:

$$y_n = \sum_k \sum_m a_{km} \cdot |x_{n-m}|^k \cdot x_{n-m}. \tag{2.9}$$

Based on what was mentioned in Section 2.3.3, even-order terms are usually used in MP models. It can also be represented as (1.9), which indicates that the model includes nonlinear memory effects. However, since it does not have cross-terms, its modelling capability can be sub-optimal in some applications [64].

Envelope-memory polynomial model

The evnelope-memory polynomial (EMP) model is given by [62]:

$$y_n = \sum_k \sum_m a_{km} \cdot |x_{n-m}|^k \cdot x_n = x_n \cdot \sum_k \sum_m a_{km} \cdot |x_{n-m}|^k. \tag{2.10}$$

Its difference as compared with MP is the complex input value in the formulation, x_n in EMP versus x_{n-m} in MP. It assumes that only the gain of the model, i.e., $\sum_k \sum_m a_{km} \cdot |x_{n-m}|^k$, has memory. Therefore it has some cross terms, and can have better performance than the MP model in some applications.

Dynamic deviation reduction model

In most of the reported baseband Volterra-based models with cross terms, the terms are of the form $|x_{n-p}|^k x_{n-q}$. However, starting from the RF representation of Volterra series (2.8), other cross terms with the same input samples and the same order can be derived. For example, both $|x_{n-p}|^k x_{n-q}$ and $|x_{n-p}|^{k-2r+2} x_{n-p}^r x_{n-q}^{*\,r-1}$ can be derived from the same term in the RF representation, but the latter is not present in most of the proposed models. Zhu in [63] used such terms to obtain an enhanced model called the dynamic deviation reduction (DDR) model, and considerable results have been reported. The first-order DDR model is given by [63]:

$$y_n = \sum_k \sum_m a_{km} \cdot |x_n|^k \cdot x_{n-m} + \sum_k \sum_m b_{km} \cdot |x_n|^{k-2} \cdot x_n^2 x_{n-m}^*. \tag{2.11}$$

Generalized memory polynomial model

The original form of memory polynomial in (3.4) can be modified to include cross terms. A generalized memory polynomial (GMP) model is proposed in [64], reporting performance improvement as compared to some other models, including the MP model. It can be represented by:

$$y_n = \sum_k \sum_m \sum_q b_{kmq} \cdot |x_{n-m-q}|^k \cdot x_{n-m}. \tag{2.12}$$

2.4 Figures of merit

It is usually required to assess the performance of a system or test the effectiveness of a technique or a method on a system. For such purposes, figures of merit (FOMs) are employed, which usually quantify the performance.

Two classes of FOMs are widely used for performance assessment of behavioral models. The first class includes the FOMs that mainly imply in-band accuracy. Two common ones are error vector magnitude (EVM) and NRMSE. They are actually equivalent to each other. However, EVM is usually expressed in per cent, whereas NRMSE is usually in decibel (dB). They are based on the error signal defined by:

$$e_n = y_n - \hat{y}_n, \tag{2.13}$$

where y_n and $\hat{y}_n$ are the real and estimated signals, respectively. EVM is given by [13]:

$$\text{EVM} = \frac{\sqrt{\frac{1}{N}\sum_{n=1}^{N}|e_n|^2}}{\sqrt{\frac{1}{N}\sum_{n=1}^{N}|y_n|^2}}. \tag{2.14}$$

In other words, it is the ratio of the root mean square of the error signal to the root mean square of the real signal. In terms of power, it is the power ratio of the error signal to the real signal.

The second class of FOMs are those concerning out-of-band accuracy. The most used FOM of this class is the adjacent-channel power ratio (ACPR) or adjacent-channel leakage ratio (ACLR), usually expressed in dB. It is the power ratio of the signal spectrum leaked to the adjacent channel due to nonlinearity in the main-band signal spectrum, i.e., [13]:

$$\text{ACPR} = \frac{\int_{f_{adj}} |Y(f)|^2 df}{\int_{f_{ch}} |Y(f)|^2 df} \tag{2.15}$$

where f_{ch} and f_{adj} are the frequencies in the main band channel and adjacent channel, respectively. $Y(f)$ is the Fourier transform of the signal.

2.5 Predistortion

One of the profound applications of behavioral models is their use in predistortion. Predistortion of a circuit/system means adding a functional block prior to the non-ideal circuit, so that the total response becomes as linear and memoryless as possible. The concept is shown in Figure 2.5 for a simple two-tone case. Without predistortion in Figure 2.5a, intermodulation (IM) products are present at the PA output. By applying predistortion prior to the PA, a predistorted signal excites the PA. This signal has two IM products such that they cancel out the IM products generated by the PA. The result is a clean two-tone signal at the output, as shown in Figure 2.5b.

The predistortion block can be a software algorithm and/or a hardware circuit. The hardware predistorter can itself be an analog circuit or a digital one. DPD is

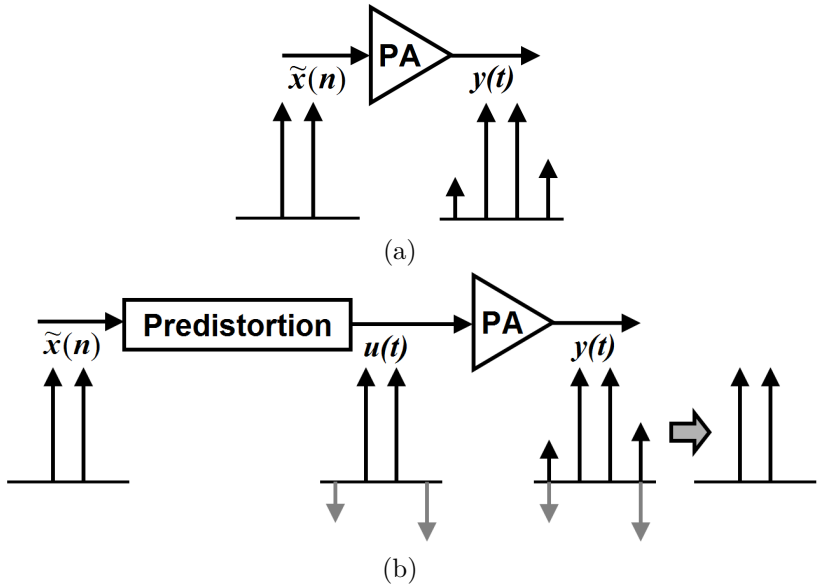


Figure 2.5: Demonstration of predistortion concept in two-tone case. The black small tones in (a), generated at the IM output due to intermodulation, are cancelled out in (b) by injecting the gray tones to the PA input. The gray tones are equal to the black ones in amplitude, but with opposite phase.

increasingly of high interest [67, 68, 69, 70] because of many reasons [71]. Digital predistorters are flexible, accurate, reliable, and persistent against changes. Moreover, digital circuits usually consume less energy than analog circuits that have active devices. They are also appropriate for adaptive predistortion, to deal with the change of circuit characteristics over time. Some types of digital predistorters are explained in this PhD thesis. They are based on digital signal processing techniques.

The DPD approach, used in this chapter and most of the other reports, is briefly illustrated in Figure 2.6. In this approach, first two samples from the input and output of the non-ideal circuit, the PA in this case, are taken. After conversion to baseband, producing $\tilde{u}(t)$ and $\tilde{y}(t)$, a behavioral model is built with $\tilde{y}(t)$ as the input and $\tilde{u}(t)$ as the output. Learning such an inverse model is called indirect learning. The learned inverse model is then copied as the predistortion block, and will predistort the input signal to have desired output. In DPD, the predistortion block is in digital domain, hence before digital-to-analog conversion, as can be seen in Figure 2.6. DPD techniques can compensate not

only nonlinearity and memory effects, but also IQ imbalance [25, 26, 27].

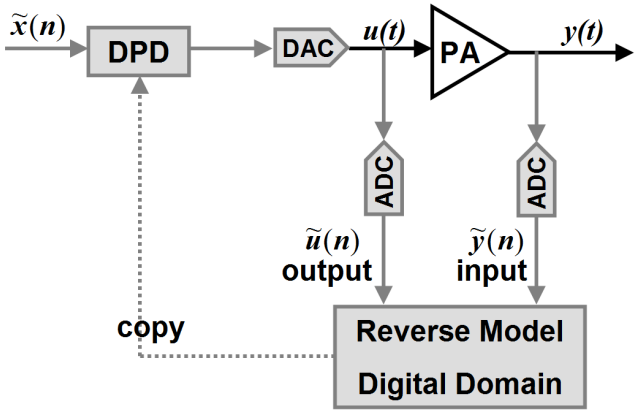


Figure 2.6: DPD approach based on indirect learning.

2.6 Conclusion

In this chapter, system-level behavioral models capable of modeling and compensating non-ideal effects are introduced, including S-functions, AM/AM and AM/PM models, and Volterra-based models. Different variations of Volterra-based behavioral models are also discussed, and two important metrics to assess the performance of behavioral models are introduced. The introduced concepts and definitions will be used in the next chapters.

Behavioral models are strong tools for system-level analysis. A profound application of such models is DPD. DPD techniques usually use Volterra-based models to cancel or reduce the non-ideal effects in RF circuits and systems. These techniques will be elaborated in the next chapters.

Chapter 3

Modeling and linearization of non-ideal circuit

The contents of this chapter is based on the author's papers presented in different conferences [1, 2, 3, 4].

3.1 Introduction

As mentioned in the previous chapter, the unwanted non-ideal effects can be removed by some techniques. Particularly, the removal of nonlinearity, called linearization, is of great interest. The term linearization is sometimes used for a more general purpose, namely for the compensation of all kinds of non-ideal effects mentioned in Chapter 2. In this chapter, some linearization techniques based on what is discussed in Chapter 2 are presented, and some applications are mentioned.

Section 3.2 presents a new formulation of S-functions, and uses that formulation for modeling and linearization of mixers. Section 3.3 presents a Volterra-based model, and Section 3.4 explains a DPD approach based on that model. An application of DPD techniques in active antenna arrays is demonstrated in Section 3.5, and finally, the chapter is summarized is Section 3.6.

3.2 S-functions model for mixers

The contents of this section is based on the paper presented in the 6^{th} European Microwave Integrated Circuits Conference (EuMIC), 2011 [1]. The PhD thesis' author contributed to this paper with the proposed formulation of S-functions for intermodulation modeling.

3.2.1 Theory

The classic form of S-functions introduced in Section 2.3.1 expresses the relationships between different frequency components located at the fundamental frequency and its harmonics. This section proposes another way of S-functions representation, so that it can be used in multisine case for expressing the relationships between the main tones and their intermodulation products. In this way, we are able to formulate the intermodulation and in-band distortion around the fundamental frequency. Therefore, the assumption here is that we have enough suppression of higher-order harmonics, so that their effects can be ignored.

We show our approach for mixer modeling. However, the approach is general and can be used for nonlinear two-port blocks. There are some articles in the literature on mixer modeling by using S-functions [72, 73]. In [72] the mixer was treated as a block with three ports; the RF port and LO port were excited with two large tones, and the reflected waves were estimated at the RF and IF ports. However, the model in [73] was more general and considered a multi-tone multi-port case. Nevertheless, in none of the [72] and [73] intermodulation products were tried to be estimated.

The modified model for mixer developed in this section describes the mixer as a two-port, black box, frequency converting device with an internal local oscillator (LO) source, as shown in Fig. 3.1. Only the in-band frequencies will be considered, assuming that the undesired out-of-band frequency components are suppressed by filters. With this hypothesis, all the Sf and Sfc functions of the model will be dependent not only on the large signal operating point, but also on the magnitude and phase of the local oscillator waveform. Providing frequency conversion, the mixer is itself a nonlinear device, but even so the conversion characteristic has to be linear to avoid distortion of the converted signal. In other words, the ideal mixer is a multiplier, or a second order nonlinear block. However, other nonlinearities also exist in real mixers, leading to generation of harmonics and intermodulation products. As mentioned above, modeling of

intermodulation products is of interest in this section. The extracted model has to predict the response of the mixer to the excitation by two tones with large amplitudes involving nonlinear operation. Even though this model has been developed for a simple two-tone modulated signal, it could be easily extended to account for few more tones.

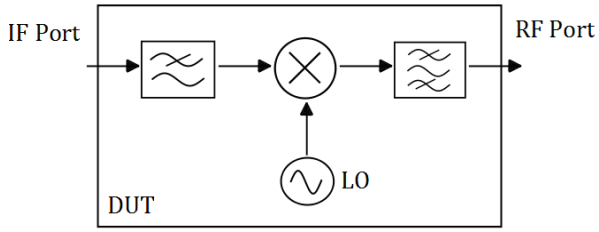


Figure 3.1: Mixer used as a two-port block for modeling [1].

To adapt the formulation of eq. (2.3) to the mixer, two main changes have been introduced. The first is that the frequency grid must fit the input and output spectra of the mixer, so the input frequency grid will be different from the output grid in some frequency shift. The second change is that the large signal operating point is no more driven by a single large tone but by two equal-amplitude large tones, that can have different contributions to the output. Thus the memory effects can be modeled by this notation.

Note that in previous works, the modulated continuous-wave (CW) signal was modeled in the baseband domain [52], assuming no memory, and the extension of the model has been developed to take into account the memory effect, using a hidden function which leads to a large amount of computation [54, 55]. Nevertheless, the proposed notation in this paper provides a simple method to model the memory effects under two-tone excitations.

In the new model the reference for the phase is the envelope of the two large tones which impose the operating point. In the native amplifier model, the reference at which all the input waveforms are aligned is the point at which the single large signal tone has zero phase. In the two-tone large signal excitation, the normalization point is the one at which both the input large tones have zero phase, in order to have the same simplifications as in eq. (2.3). In this configuration, the input frequencies are not a multiplication of each other. Therefore, it is necessary to find the delay τ to be applied to all the inputs for alignment. This can be found in a closed form by some mathematical calculations provided in Appendix A. For example, in its simplest form, when the greatest common divisor (GCD) of the tone frequencies equals their difference, τ can be

found by:

$$\tau = \frac{\phi\left(A_{1,1}\right) - \phi\left(A_{1,-1}\right)}{2\pi\left(f_{1,1} - f_{1,-1}\right)}, \tag{3.1}$$

where the notation from Fig. 3.2 has been adopted. The delay calculation method for more general cases is explained in Appendix A. With this delay value, the model of eq. (2.3) can be re-written as:

$$\begin{aligned} B_{pm} = &\sum_q \sum_n Sf_{pmqn}(|A_{11}|) \cdot A_{qn} \cdot \exp\left(j2\pi(f_m - f_n)\tau\right) \\ &+ \sum_q \sum_n Sfc_{pmqn}(|A_{11}|) \cdot A^*_{qn} \cdot \exp\left(j2\pi(f_m + f_n)\tau\right), \end{aligned} \tag{3.2}$$

where the meaning of the symbols are the same as in eq. (2.3). The dependency of the Sf and Sfc functions on the local oscillator wave is not considered, because its contribution is a same shift in the phase of all the frequency components which has no effect in the linearization procedure.

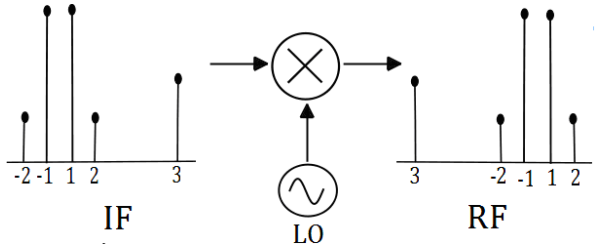


Figure 3.2: IF and RF frequency spectra, and the way the tones are numbered in eq. (3.2) [1].

3.2.2 Model identification

To verify the methodology, a simulation based extraction was performed. The device-under-test is the circuit model of a doubly balanced diode mixer by which the diodes are placed in the ring configuration. All the simulations were performed in the advanced design system (ADS) environment. In this section, only the third-order intermodulation products are considered. However, the technique can be easily employed to include higher-order intermodulation products.

The mixer is set-up as an up-converter. The two tones of the large signal have a frequency spacing of 2 MHz and are centered at 300 MHz, the LO is at 2.7

GHz, and the output spectrum is thus centered at 3 GHz. The frequencies at which the tickle are applied during the model extraction are the ones shown in Fig. 3.2. There are five tones in total per port: the two large tones are denoted with indexes (-1,1) , the frequencies at which the 3^{rd} order intermodulation products lie with (-2,2) and the LO with (3). This last tickle is added to include the possible leakage of LO signal at the intermediate frequency (IF) and RF ports. At the output, the same pattern shifted in frequency is evaluated.

For model extraction, a set of harmonic balance simulations were performed. The device is driven by a fixed two-tone large signal, and then a tickle is in sequence applied to every frequency point of interest. For the extraction of the Sf and Sfc functions the phase shifting technique [53] was used. This set-up is repeated for a number of large signal operating points: in our case, a sweep with 2 dB step was made from -5 to 19 dBm.

All the calculations are performed in Matlab and the values of the model are then saved in a citi-file. This citi-file is used to define the equations of a frequency-defined device (FDD) component in the ADS schematic, which permits the definition of behavioral equations of the device and can be excited by input voltages and currents applied in a typical schematic design. In the FDD component, a frequency list of only the characterized frequencies was created in order to index voltages and currents, automatically turning to zero the output at any other frequency. When inserted in the FDD component, the values of the S-functions at different operating points are linearly interpolated.

3.2.3 Results

To have the values for the model validation, further sets of simulations are performed on the circuit model of the doubly balanced diode mixer. These simulations are set-up at large signal operating points different from the points at which the model was extracted. In this way we are sure that the validation data is different from the data used for identification. In the particular case, a sweep from -4 to 18 dBm with 2 dB step was done in the intermediate points that in the FDD model are linearly interpolated. A schematic harmonic balance simulation which includes the FDD component is set up in the microwave circuit simulator. The outputs predicted by the S-functions model are then compared with the output of the circuit model. Fig. 3.3 and Fig. 3.4 show the consistency of the predicted response of the proposed model with that of ADS in magnitude and phase in terms of errors, respectively. The increasing contributions of Sf and Sfc functions to the output signal when the mixer is driven to increasing

compression conditions are shown in Fig. 3.5 with the notation introduced in Fig. 3.2.

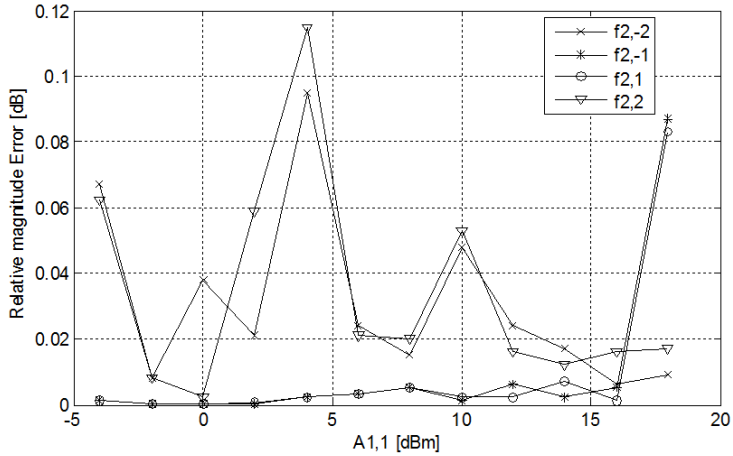


Figure 3.3: Relative magnitude error for the S-functions model [1].

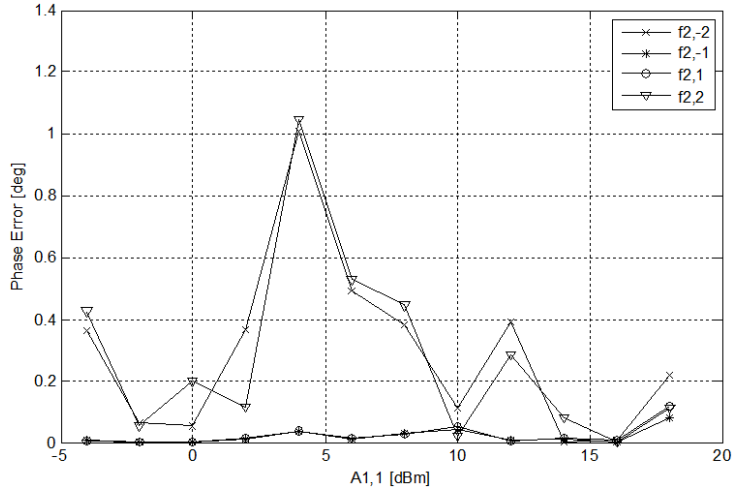


Figure 3.4: Phase error for the S-functions model [1].

Since the FDD implements only the equations in (4), the simulation time is reduced almost five times, maintaining the same reliability of the results. For

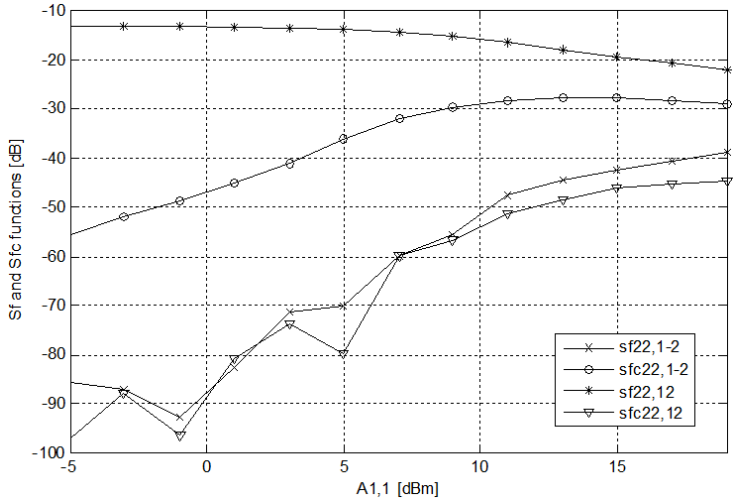


Figure 3.5: Sf and Sfc functions for the mixer [1].

instance the simulation time in 1 dB compression point is 0.75 sec for S-functions model, whereas it is 3.55 sec for the circuit model.

3.2.4 Linearization

The extracted S-functions model provides a reliable prediction of the circuit response. Moreover, it can be used for predistortion purposes to suppress spurious products, especially the 3rd order intermodulation products, which are very close to the signals of interest and very hard to filter out.

In this work the predistortion by signal injection has been used. In particular, some other signals are injected at the input in order to suppress the intermodulation products by phase compensation. In our simulation set-up based on the extracted S-functions model, the two frequency components were injected at the location of third order intermodulation products, namely 297 MHz and 303 MHz. The signal to be injected can be calculated from eq. (3.2). In fact, it has been tested that only one or two, most dominant S-functions coefficients are sufficient to maintain the linearization performance. These coefficients are selected by the influence of each single coefficient to the output. In the example considered in Fig. 3.5, the most dominant coefficients are $Sf_{22,12}$ and $Sfc_{22,1-2}$. As a result, the first order approximation of the predistorsion

signal at 303 MHz is given by

$$A_{12} = -\frac{B_{22}}{Sf_{22,12}\left(|A_{11}|\right)} \tag{3.3}$$

where the output phasor B_{22} is obtained from eq. (3.2) when only two-tone large signal excitation is applied to the model. It should be highlighted that the magnitude of the injected predistorting signal has to be small enough to satisfy the S-functions model assumptions. To illustrate this we plot in Fig. 3.6 and Fig. 3.7 the results of the linearization at 1 dB and 3 dB compression points, respectively for increasing values of the injected signals, and compare the output of the S-functions model with the output of the circuit model. For mild nonlinear operating point, as in Fig. 3.6, the two models are consistent. When the device is driven into higher nonlinear conditions, as in Fig. 3.7, the linear interpolation of the S-functions values is less accurate and the superposition principle approximation is valid for relative smaller inputs. Note that the phase relationship between the two large tones is not influenced by the introduction of the injected signal at the input as long as the tickles are small.

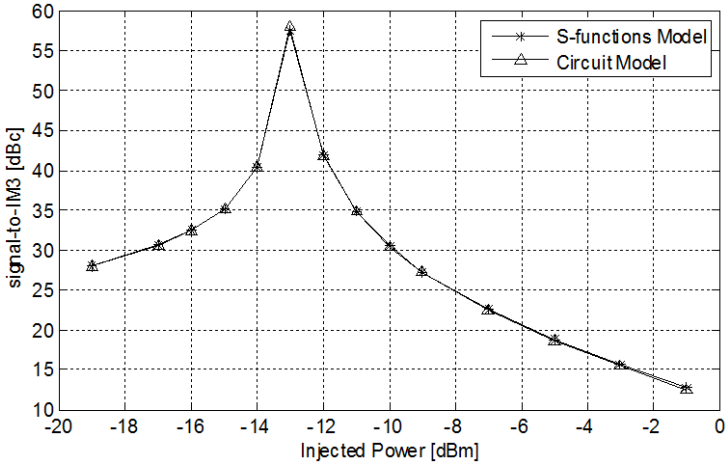


Figure 3.6: Linearization for 1 dB compression [1].

3.3 Cross-memory polynomial model

The contents of this section is based on the author's paper presented in the 7^{th} European Microwave Integrated Circuits Conference (EuMIC), 2012 [2].

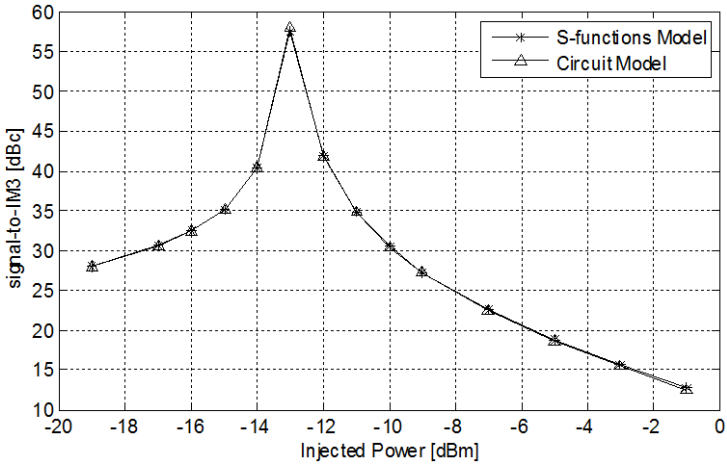


Figure 3.7: Linearization for 3 dB compression [1].

In the previous section, linearization was done by using S-functions. However, the most widely used models for linearization are Volterra-based models. This section introduces a novel Volterra-based model which will be used for linearization in the next section.

3.3.1 Theory

As mentioned in Section 2.3.3, a conventional MP model is formulated as [61]

$$y_n = \sum_k \sum_m a_{km} \cdot |x_{n-m}|^k \cdot x_{n-m}. \tag{3.4}$$

where k, m and a_{km} are the nonlinearity order, the memory depth and the weight of each time sample x_{n-m}, respectively.

The MP model yields very good results in modeling nonlinear circuits with memory. As stated earlier in Section 2.3.3, its performance can even be enhanced if the cross-terms are included in the formulation [64, 8]. The modified model can be, for example, the generalized memory polynomial, introduced in Section 2.3.3.

This section proposes a model [2], which has the simplicity of the memory polynomial formulation, but partly benefits from the cross terms, like in GMP. The idea is to use the magnitude of the present input sample to compute the

nonlinear weights of the memory terms. This modification is originated from the fact that the system has memory, and its recent samples are linked together. Evaluating the validity of this assumption is the subject of this section. The suggested new model is written as

$$y_n = \sum_{m=0}^{M-1} g_m\left(x_n\right) x_{n-m} \tag{3.5}$$

$$g_m\left(x_n\right) = \sum_{k=1}^{K} a_{km} |x_n|^{k-1} \tag{3.6}$$

The number of coefficients in this formulation is the same as that of MP. It should be mentioned that, although only the first order of the delayed input is used in the formulation, the method still models nonlinear memory effects. That is because the weight of each term is still nonlinear, even though it is now a nonlinear function of only the current sample. Like the MP model, the equations for determining the coefficients are linear, and common LS methods like recursive least square (RLS) can be used to find them [74].

Since the memory term is at a different time from the variable of the nonlinear weight for $m \neq 0$, eq. (3.5) has cross terms which endow the model with an improved characterization performance compared to those found in the GMP model, and which can compensate for the simplification of the nonlinear function in eq. (3.6). This feature is discussed in the next subsection. Moreover, the absolute input value only at the present time is considered for the nonlinear gain computation. Hence the complexity, especially for field-programmable gate array (FPGA) implementation, can be reduced and the model is more computationally and energy efficient although the number of the coefficients is the same as the memory polynomial. This gain of efficiency can be illustrated by referring to table 3.3.1 [8] and Fig. 3.8. In all diagrams of Fig. 3.8, the complex-complex multipliers are illustrated in dark gray, while the light gray circles are complex-real multipliers and the white circles indicate real-real multipliers. If a standard architecture for the FPGA implementation is used as in Fig. 3.8a, more floating-point operation per second (FLOPS) will be utilized than if the proposed architecture in Fig. 3.8b is used. This is mainly due to the higher number of complex-complex multiplications in Fig. 3.8a. It is worth mentioning that the propagation delay difference due to the branch in Fig. 3.8b connecting the input to the last multiplier is not a critical issue and can be compensated, if required, by inserting a delay block in that branch.

If the same optimized architecture is applied in CMP, it can be easily seen that a lower number of delay blocks than in MP is required, while the rest of the

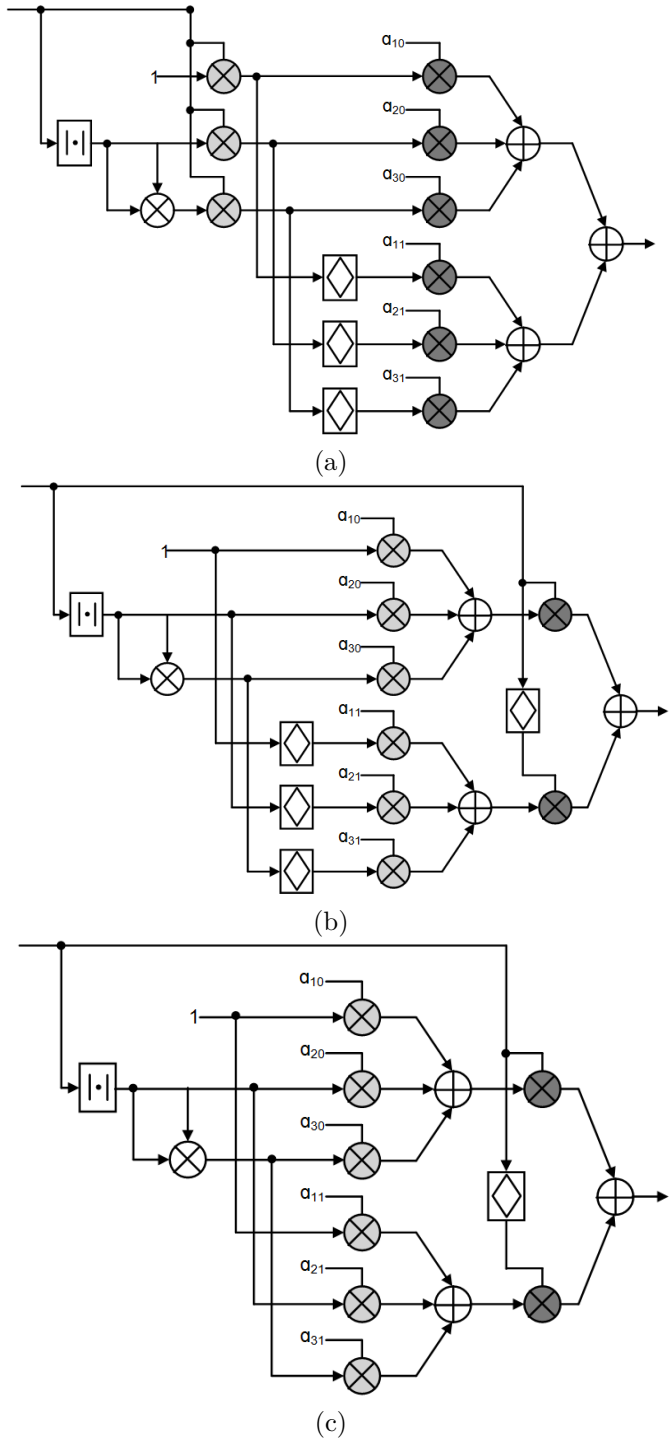


Figure 3.8: Standard architecture of MP, b) proposed architecture for MP, and c) proposed architecture for CMP [2].

Table 3.1: Number of FLOPs for different operations [8]

Operation	Number of FLOPs
Conjugate	0
Delay	0
Real addition	1
Real multiplication	1
Complex addition	2
Complex-real multiplication	2
$\|.\|^2$	3
Complex – complex multiplication	6
Square-root	6~8

circuit remains the same as for MP, as shown in Fig. 3.8c. Therefore the overall FPGA architecture complexity has been reduced by using the CMP.

3.3.2 Experimental results and discussion

The CMP model is evaluated by comparing its performance with the MP model, as well as with a memoryless model which is defined by $M = 1$ in eq. (3.5). The model has been identified by applying a wideband code division multiple access (WCDMA) single-carrier signal to the amplifier. Only 3000 time samples were used for the parameters estimation. The performance is then validated using a much longer sequence comprising about 50000 samples. The carrier frequency in both cases is 2.14 GHz. The recursive least square method is used for solving the set of linear equations to find the coefficients of eq. (3.6). The performances of the models are examined by testing two amplifier evaluation boards.

The first device under test, referred to as Amp #1, is a Sirenza amplifier evaluation board, and the second one, hereafter indicated by Amp #2, is a Freescale amplifier. Both devices are tested such that their instantaneous power exceeds the 1 dB compression point. The test setup is illustrated in Fig 3.9. It consists of an arbitrary waveform generator (AWG) and a vector signal analyzer (VSA), both connected to a PC. The baseband digital signal is generated in the PC and uploaded into the AWG. Then the AWG converts the signal into an RF time-domain waveform with a 2.14 GHz carrier frequency. The generated

AWG output signal passes through the amplifier and is detected and captured by the VSA. The VSA converts the waveform back into the baseband digital domain using its internal analog-to-digital converters (ADCs). The digitized output is fed back to the PC for processing in MATLAB. Note that in these experimental evaluations the aim is to show the suitable accuracy of the model. So a PC has been used for the computations and the FPGA architectures have not been implemented yet.

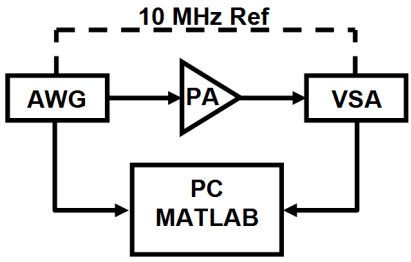


Figure 3.9: Measurement set-up for CMP modeling.

The measurement and its corresponding modeling results for a single-carrier WCDMA signal applied to Amp #1 are shown in Fig. 3.10. In this evaluation, CMP and MP have the same memory depth and polynomial order of 4 and 7, respectively. The figure illustrates the highly suitable performance of the proposed models although the instantaneous power level of the amplifier is reaching the 2 dB compression point.

To show the consistency of the CMP, the Amp #2 output has been measured and modeled. The results are shown in Fig 3.11. In this case both the MP and CMP have the same memory depth and polynomial order of 3 and 9, respectively. The convincing performance of the CMP is also proved in this evaluation which uses a WCDMA signal reaching the 2 dB compression point.

The assessment of the performance can be made easier with table 3.3.2. The criteria in the table for assessment is the NRMSE, defined in Section 2.4.

A quantitative comparison is shown in table 3.3.2 where ML stands for the memoryless model. According to the table, the CMP has a slightly better performance than MP. Note that the big differences between NRMSE values of both MP and CMP, and the NRMSE value of the memoryless model indicate the existence of memory.

The modeling capabilities of both MP and CMP have upper limit for high output

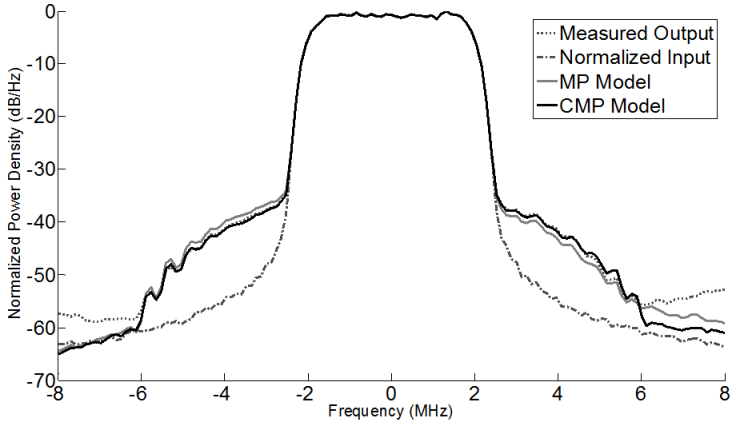


Figure 3.10: Modeling performances of MP and CMP for Amp #1 with a measured 1xWCDMA signal reaching the 2 dB compression point of the DUT [2].

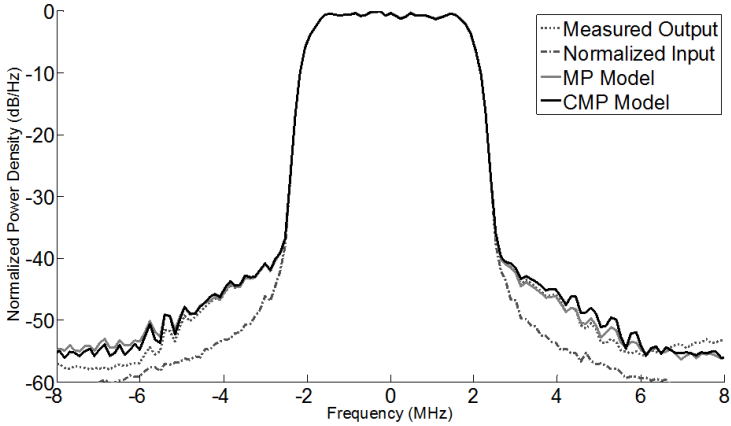


Figure 3.11: Modeling performances of MP and CMP for Amp #2 with a measured 1xWCDMA signal reaching the 2 dB compression point of the DUT [2].

Table 3.2: Performance of CMP and MP models for measured WCDMA signal reaching the 2 dB compression point [2]

		NMSE (dB)
Amp #1	ML	-27.4
	MPM	-38.5
	CMPM	-39.1
Amp #2	ML	-27.8
	MPM	-39.1
	CMPM	-39.6

power. This fact can be seen when the output power of Amp #2 for instance is increased to reach the 2.5 dB compression point. Although both models in this case still have NRMSE values higher than -39 dB, their predictions of adjacent channel power levels start to degrade as shown in Fig. 3.12. Therefore, like any other models care has to be taken when working with the CMP and MP at power levels close to saturation.

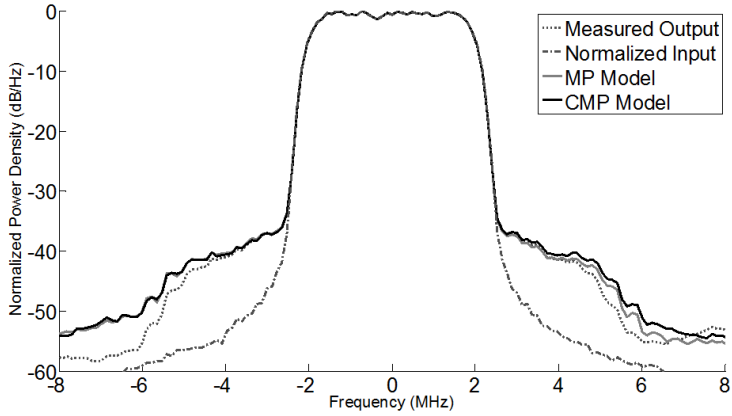


Figure 3.12: Modeling performances of MP and CMP for Amp #2 with a measured 1xWCDMA signal reaching the 2.5 dB compression point of the DUT [2].

All these results confirm the reliable performance of CMP, although the nonlinear weight is simplified into a function of only the present input sample. The model has the capability to model the behavior of a microwave power amplifier with power levels close to saturation.

3.4 Linearization with CMP model

The contents of this section is based on a paper presented in International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits (INMMIC), 2014 [3]. The work was supervised by the PhD thesis' author.

The previous section presented the CMP model for PA modeling. This section uses the CMP model for PA linearization. The proposed technique belongs to the DPD class, discussed in Section 2.5. Therefore, the discussion in this section is in the digital baseband domain.

3.4.1 Experimental setup

The experimental hardware setup used in this study is shown in Fig. 3.13. It consists of an FPGA board, a data converter board, lowpass filters to suppress unwanted spectrum, up- and down-converters, power splitter for splitting the LO signal and a power amplifier to be linearized. For FPGA, a Xilinx ML605 FPGA design kit was used. The ML605 board hosts an FMC150 daughter board containing a 14-bit ADC and a 16-bit digital-to-analog converter (DAC) converters operating at 245.76 MSamples/s. The FMC150 board has a low frequency limitation of 3 MHz at the MMCX interfaces. Therefore, an IF stage is mandatory. The architecture of the DPD-based linearizer is presented in Fig. 3.14.

The baseband stage is divided into two parts. The first part is implemented in a PC. It consists of a MATLAB script which reads a data set of excitation signals (e.g., single- and two-carrier WCDMA). The excitation signal is multiplied with the parameter A_{CP}, which adjusts the signal amplitude in order to reach the PA gain compression point of interest. The MATLAB script is executed twice. In the first run the excitation is transmitted, without performing DPD, throughout the loop and is captured back on the PC side. Then, the received signal is corrected by a gain factor $1/G$ such that both the transmitted and received signals have the same average power. The DPD training is finally performed using the indirect learning technique presented in [75]. The model coefficients are estimated using the RLS method. In the second run, the excitation signal is predistorted and transmitted once again. The transmitted IQ signal samples have 16 bits each in order to fit the DAC input length. The second part is implemented in the ML605 FPGA board. The communication between both parts is realized through a universal asynchronous receiver/transmitter (UART) interface at a baudrate of 115.2 kbps. Apart from the UART controllers,

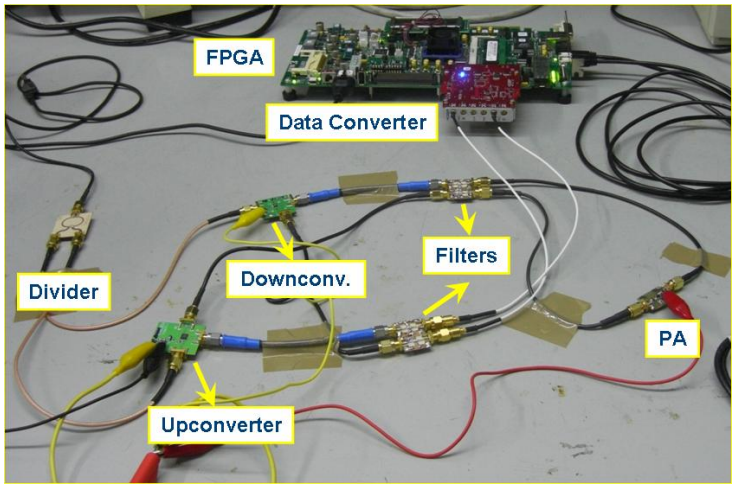


Figure 3.13: FPGA-based DPD setup.

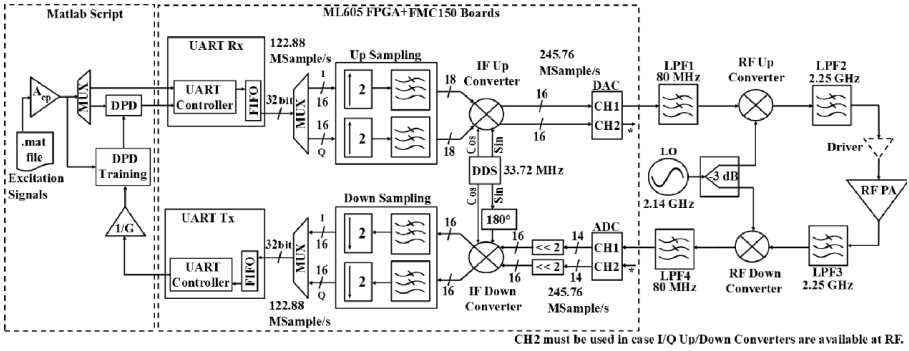


Figure 3.14: FPGA-based DPD architecture [3].

UART Rx and Tx also contain a 128 kB first-in first-out (FIFO) block for IQ signal components, which stores up to 64 kSamples. A clock signal running at 245.76 MHz is distributed to the ADC/DAC on the FMC150 board as well as to the ML605 FPGA board as to avoid synchronization issues. However, a high-speed clock signal may cause timing issues if the DPD is fully implemented on the FPGA itself. The clock speed should not be too low either such that the Nyquist bandwidth accommodates the IM products generated by DPD. Hence, a trade-off is established by using two sampling rate domains. One at 245.76 MSamples/s for the ADC/DAC and another at 122.88 MSamples/s for signal processing. Thus, allowing a wide bandwidth of 61.44 MHz, which covers up to the 5^{th} order IM products of single- and two-carrier WCDMA signals and 3^{rd} order of three- and four-carrier WCDMA signals. The intermediate frequency (IF) LO is set to 30.72 MHz, which places the predistorted signal at the center of the 61.44 MHz band, plus a 3 MHz shift due to the low frequency limitation imposed by the FMC150 board, resulting in 33.72 MHz. IF LO is generated with a direct digital synthesizer (DDS) and the IF up- and down-converters use complex multipliers. In order to comply with the WCDMA standard, RF LO is set to 2.14 GHz using a RF signal generator. Finally, high input third-order intercept point (IIP3) RF up- and down-converters were used in order to guarantee the PA is the main source of nonlinearity present in the loop.

3.4.2 Experimental results

In order to assess the effectiveness of the CMP for DPD as well as to verify the functionality of the proposed DPD linearizer architecture, two experiments were set up. Firstly, a wideband driver amplifier by Avago Technologies (ABA-54563), hereafter referred to as Amp #1, was evaluated under the excitation of a single-carrier WCDMA signal reaching its 2 dB compression point.

Figure 3.15 presents the measured AM/AM performance for the cases where no DPD is performed, DPD with memoryless (ML), DPD with MP and DPD with cmp are performed. For MP and CMP DPD, $K = 5$ and $M = 3$. The performance of CMP DPD is clearly very similar to that of MP DPD. Figure 3.16 shows a 12 dB sideband suppression for MP and CMP DPD as well as 8 dB for ML DPD as compared to the performance in the absence of DPD.

The second experiment was carried out cascading the driver amplifier Amp #1 with a GaAs InGaP HBT PA by Hittite Microwave (HMC413) under the excitation of a two-carrier WCDMA signal reaching its 1.5 dB compression point. The cascade arrangement will be referred to as Amp #2. Similar to the experiment with Amp #1, $K = 5$ and $M = 3$ for MP and CMP DPD. The

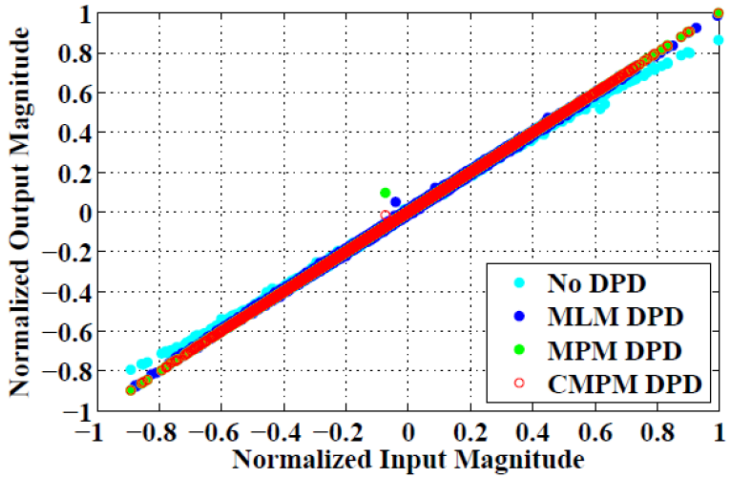


Figure 3.15: Measured AM/AM performance of Amp #1: single-carrier WCDMA signal reaching the 2 dB compression point. [3].

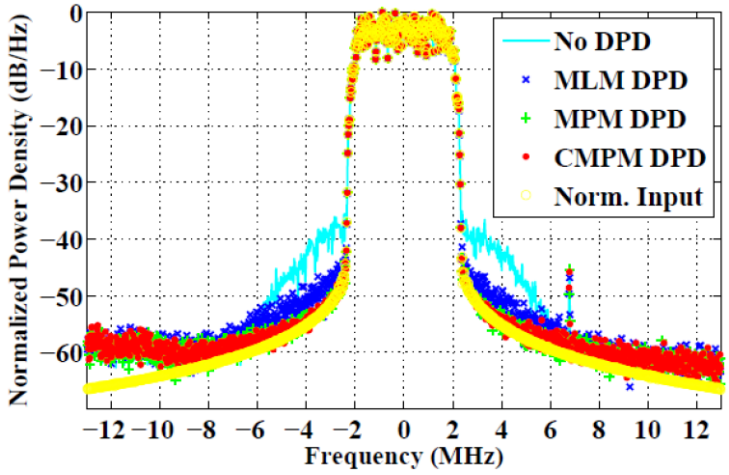


Figure 3.16: Measured output spectra of Amp #1: single-carrier WCDMA signal reaching the 2 dB compression point. [3].

AM/AM performance is presented in Fig. 3.17, in which one can observe that the memory effects are more pronounced. This is due to the fact that a wider bandwidth signal is applied as excitation.

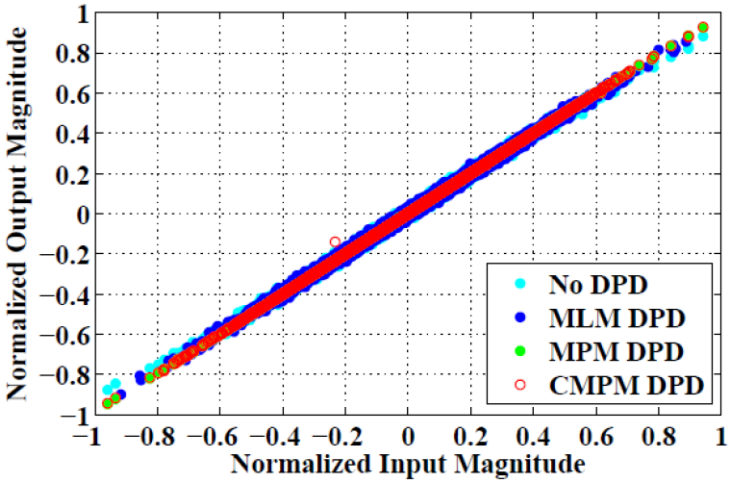


Figure 3.17: Measured AM/AM performance of Amp #2: two-carrier WCDMA signal reaching the 1.5 dB compression point. [3].

Once again, the MP and CMP DPD present very similar performances where an 8 dB sideband suppression is achieved as shown in Fig. 3.18. However, as expected the performance of the ML DPD in suppressing spectral regrowth is now very poor. A trivial 2 dB suppression is achieved. Finally, the NRMSE and ACPR metrics were evaluated as to give a more quantitative grasp on the similarity of MP and CMP DPD performances, as highlighted in Table 3.3. A maximum of 0.8 dB difference is observed comparing those metrics.

3.5 Application example: active antenna array

The contents of this section is based on the author's papers presented in IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR), 2013 [4].

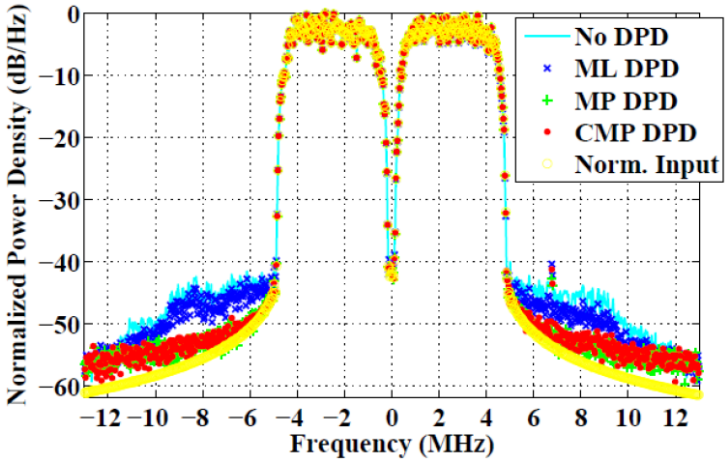


Figure 3.18: Measured output spectra of Amp #2: two-carrier WCDMA signal reaching the 1.5 dB compression point [3].

Table 3.3: NRMSE and ACPR performances of Amp #1 (excitation reaching 2 dB CP) and Amp #2 (excitation reaching 1.5 dB CP) [3].

	DPD Model	NMSE (dB)	ACPR (dBc) @ +5MHz Offset
Amp #1 (single-carrier WCDMA)	No DPD	-29.5	-43.3
	MLM	-31.6	-43.9
	MPM	**-45.6**	**-46.0**
	CMPM	**-46.4**	**-45.3**
Amp #2 (two-carrier WCDMA)	No DPD	-24.5	-40.4
	MLM	-24.7	-43.0
	MPM	**-43.4**	**-45.5**
	CMPM	**-42.9**	**-45.7**

3.5.1 Problem statement

The adoption of active antenna array (AAA) or smart antennas have the potential to significantly enhance the capacity and efficient use of spectrum and power in wireless communication systems compared to the existing methods. Although the underlying principles behind the technology are well established and have been used in radar and sonar applications for many years, it is only now that the processing power and antenna technology has reached a point where it is becoming commercially viable. With this change in application area there is a corresponding shift from high to low cost production. This shift has implications for component tolerances and manufacturing quality for the basestation. As a result there are expected to be variations in the characteristic performance of each transceiver chain. Furthermore, AAA basestations are designed with the intention of having all electronic components housed in a single enclosure and mounted at a height directly behind the radiating elements. If one PA of, for instance, a 4-by-4 array were to fail and needed to be replaced, it is not feasible to have a replacement that can exactly match the characteristics of the other 15 [76]. Moreover, there will be a temperature difference even within the AAA enclosure since the preference is that these systems should use passive cooling techniques. Thus, heat energy will not have an even distribution inside the enclosure which leads to uneven temperature distribution throughout the AAA, resulting in some variations in amplifier responses. Hence, there are many sources of variations in the amplifiers of an AAA. Therefore, a solution is needed to provide first the ability to model and then the capability to compensate for variations in the multiple signal paths such that the resulting beam pattern is transmitted as intended and the AAA is not adversely affected by the variation in power amplifiers or by replacing power amplifiers.

Taken individually, each PA can be characterized to a suitable accuracy by an MP model [61, 64]. This class of models has proved to have high capabilities in accurate modeling of nonlinear circuits with nonlinear memory effects [64, 8].

This section proposes an approach to efficiently model an array of amplifiers. It is composed of a core model along with differential models as many as the amplifiers. Due to the simplifications made to the model, it uses a very compact set of parameters which highly simplifies the training phase of modeling the complete array. This approach can also be used in linearization to reduce the total number of parameters, resulting in a much simpler training phase for the whole array.

3.5.2 Modeling approach

In this section we investigate two modeling approaches for active antenna arrays. One approach is to use identical memory polynomial models [61] for all PAs. The second is to provide one core memory polynomial model [61] in common for all the PAs, along with differential blocks for each PA. In this way the relative accuracy of the two approaches and the relative size for the models can be determined for a typical scenario.

In an m-by-n antenna array of identical active elements, there are $m \times n$ amplifiers of which may have variations in their characteristics, as discussed in previous section. If identical models are used for all the amplifiers, the predicted outputs cannot be modeled well due to the differences discussed previously. Hence, there is an inevitable need for modeling each amplifier separately. On the other hand, if totally separate models are used for this purpose, the number of model parameters becomes large and it can make the approach inefficient. Thanks to the similarity of the array amplifiers, some simplifications can be applied which lead to considerably lower number of parameters. The approaches are investigated using a 2-by-2 amplifier array as an example. The same strategy can be applied for a higher number of array elements. The models used in this section are all in the complex envelope domain. Note that crosstalk effects among the amplifiers are beyond the scope of this section.

Using identical models for all amplifiers

Figure 3.19 shows identical models used for modeling the four amplifiers. In this method the inputs of the amplifiers encounter the same model. This means that if the inputs are identical, the predicted outputs of the amplifiers will be the same, though it is not true in practice due to inherent differences between the amplifiers. These differences lead to considerable errors in array modeling, as will be shown later in the experimental results.

Using common-differential array model

Figure 3.20 shows the proposed common-differential array model (CDAM). It is composed of a core model (CM) which is common in all the amplifiers, and four differential models (DMs) corresponding to four PAs which model the differences between the amplifiers. Since the DM model the differential behavior of the array, namely that part of the behavior which is not modeled by the

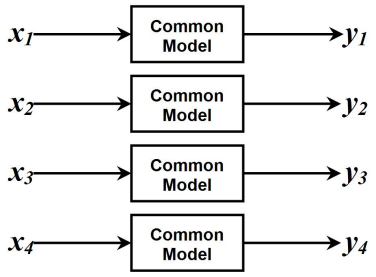


Figure 3.19: Array of four amplifiers modeled with a common model [4].

CM, their complexity is less than that of the common model. Therefore the overall array model constructed from using this approach is simpler than an array model made out of four identical models.

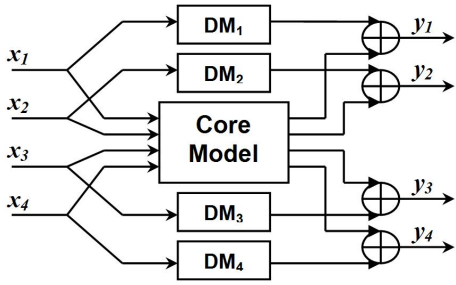


Figure 3.20: Array of four amplifiers modeled with CDAM [4].

In the next subsection, the experimental results of the two modeling approaches are demonstrated and discussed.

3.5.3 Experimental results

The four amplifiers used in this study are class AB power amplifiers using ATF52189 from Avago Technologies. These power amplifiers have been used in a testbed where the calibration of the separate AAA paths has been demonstrated [77]. For this reason the primary focus of the section is not the calibration of the antennas but the modeling of the differences between power amplifier responses. To this end a testbench using Rohde & Schwarz SMU 200A and FSQ is used to generate the PA input signal and to capture the

Table 3.4: NRMSE values for different amplifiers with CDAM

Amplifier No.	NRMSE (dB)	
	Identical Models	CDAM
1	-14.0	-32.9
2	-22.1	-37.2
3	-22.5	-31.7
4	-36.1	-36.1

output signals respectively. An identical input signal dataset of a single-channel WCDMA waveform at 2.4 GHz is used to measure each PA output response. The instantaneous output power reaches the 2 dB compression point of the PAs. The four complex envelope responses are used to extract the coefficients for both modeling approaches.

Figure 3.21 illustrates the limited predicting ability of the first modeling approach to predict the outputs. In this figure, the predicted outputs by the models are plotted against the measured outputs. In these sets of measured data, the same sample sequence of modulated signal with the same power level has been used for all amplifiers. However since they are measured at different times, the inputs are not necessarily synchronized and identical. The MP model of the fourth amplifier has been used as the common model. A quantitative evaluation of the model is shown in Table 3.4. It illustrates NRMSE values for the two approaches. As can be seen in Fig. 3.21, Fig. 3.22 and also in Table 3.4, in the first approach the modeled output does not properly match the measured output for the other three PAs. In this particular case, a big difference between the amplifiers is the latency of the amplifiers resulting in large scattering of the data in Fig. 3.21. Nevertheless, it is not only this latency which leads to the differences, but also other factors such as those addressed in Section 3.5.1, have considerable effects. Figure 3.22 is a proof of this fact, since if it was only the effect of latency, no differences would be seen in the spectrum plot. For instance, the plots for amplifier #1 in Fig. 3.22 show some mismatches between the target output and the predicted output calculated from the common model. This mismatch is not a simple shift caused by a simple gain difference and is frequency-dependent. This difference cannot be explained by a simple latency and/or gain difference.

A certain amount of spread can be seen at the upper end of the plots in Figures 3.21 and 3.23. It is mainly because the output power reaches the high 2 dB compression point and modeling is difficult in this region.

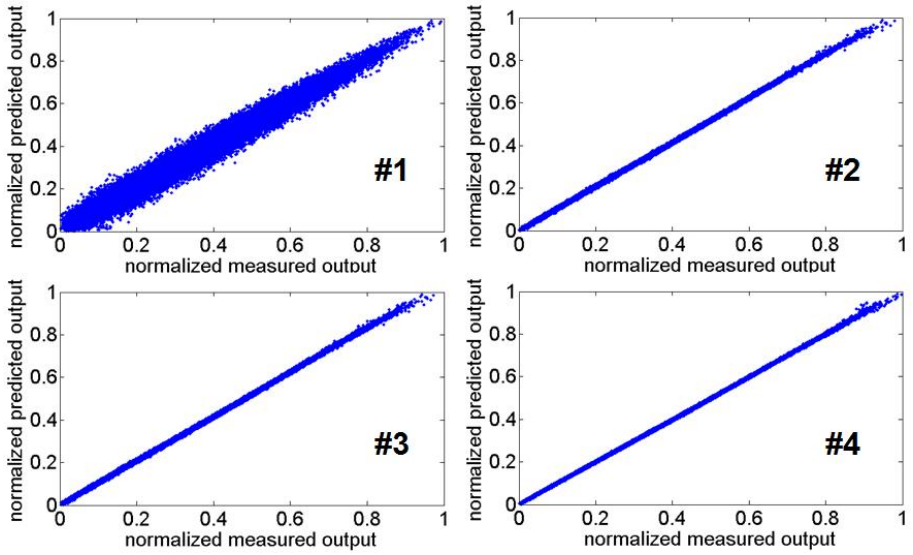


Figure 3.21: Regression curves of an array of four amplifiers modeled with a common model [4].

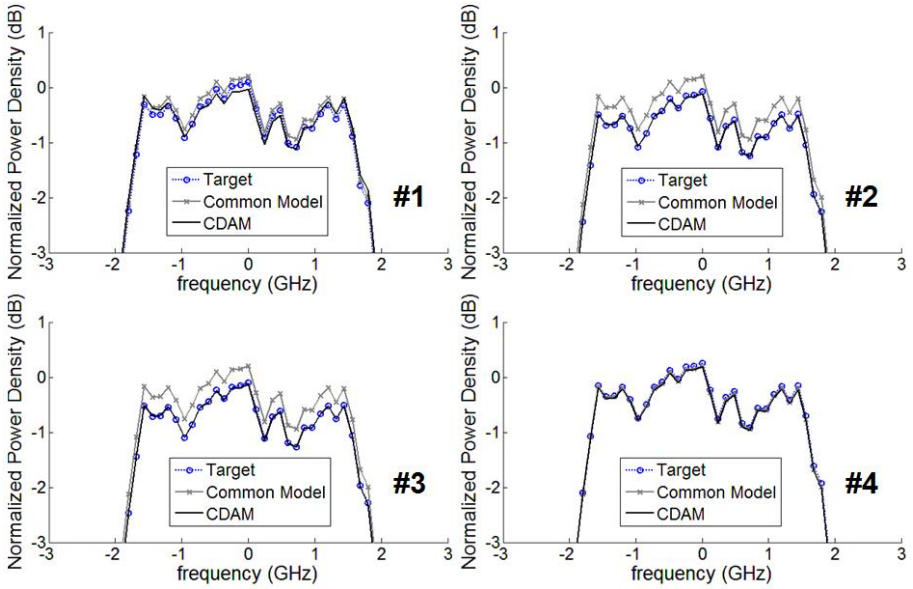


Figure 3.22: Spectrum plots of an array of four amplifiers modeled with CDAM [4].

In Fig. 3.20, as regarding the second modeling approach, the utilized core model is a memory polynomial with nonlinearity order of 6 and memory depth of 3. The differential model is a linear model with memory depth of 2. Using this linear differential model with low memory depth, remarkable improvement can be achieved. Note that the differential mode is not necessarily linear, and can be a higher-order nonlinear block, depending on the amount of nonlinearity and the amount of variations among the amplifiers. However in this special case a linear differential model is sufficient for desired accuracy.

The results are shown in Fig. 3.22 and Fig. 3.23. They show highly acceptable performance of the proposed CDAM. The regression plots of Fig. 3.23 have less scattered points and are more linear as compared to those in Fig. 3.21. Note that since the error of using the common model is negligible in out-of-band response, it is not shown in Fig. 3.22, and the figure is zoomed along the y-axis on the main band. The main variances in the amplifier responses are seen in the in-band spectrum, as in Fig. 3.22. In this figure, it can be seen that the CDAM highly outperforms the common model, and is able to compensate the variations very well. A quantitative evaluation is shown in Table 3.4. This table proves the accuracy of the proposed method. It is worth mentioning that the total number of parameters for the model in Fig. 3.20 is 34, while if four separate models were to be used the total number would be 72. The difference will be more pronounced if a 4-by-4 array is considered.

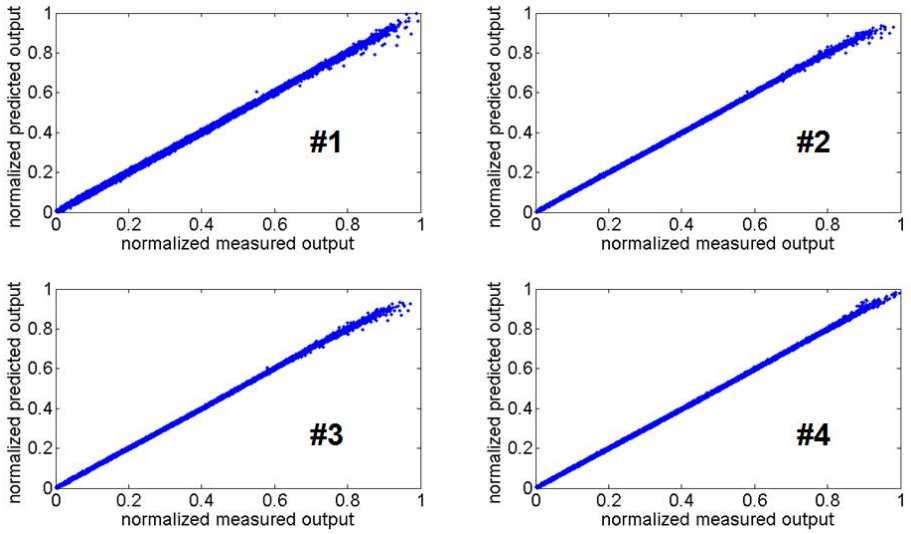


Figure 3.23: Regression curves of an array of four amplifiers modeled with CDAM [4].

3.6 Conclusions

In this chapter, different linearization techniques have been presented. The first application example was the linearization of a mixer, by using S-functions. An S-functions model was first derived, and based on the predicted value of the model, some small signals were injected to the input. The appropriate injection resulted in the cancellation of intermodulation products, and thus the linearization of the mixer.

A Volterra-based model was also proposed, which showed the estimation performance comparable with that of the memory polynomial model. Then the suitability of the model was demonstrated by applying it to an FPGA-based DPD system. This system successfully linearized power amplifiers.

Finally, a modeling approach was applied to compensate the non-ideal and non-identical effects of an array of amplifiers used in active antenna arrays. It was shown that by using a suitable simplified model, one is able to efficiently linearize the AAA system.

Chapter 4

Offset Multisine Excitation for Non-Ideality Characterization

The contents of this chapter is based on the author's paper published in the IEEE Transactions on Microwave Theory and Techniques, March 2014 [5].

4.1 Introduction

Owing to the periodicity and controllable waveform shapes of multisines, this type of excitation is a frequent choice to investigate the behavior of RF circuits and systems in response to complex modulations. Various approaches have been reported on the design of multisine signals by manipulating the amplitude and phase of their tones [78, 49, 79, 80], in order to simulate a real-world modulated signal. However, another degree of freedom corresponds to manipulating the tone frequencies, to be able to detect different nonlinearity orders in the frequency domain individually. To the best knowledge of the author, there were two proposed approaches based on this manipulation. One approach is to use a number of tones over a wide range of frequencies, and to turn off some of these tones to let higher order intermodulation products, which are smaller than the main tones, present themselves in the response [81]. This technique

has been extensively adopted [82, 83]. However, these papers only determined whether the nonlinear intermodulation product is even-order or odd-order, and did not discuss finding the nonlinearity degree of the different orders. The second approach presented in [84] focused on the second order nonlinearity only, and tried to distinguish between different second order products by engineering the frequency locations of the tones so that each of the products lies at a unique frequency. This technique was again used for a wide frequency range to characterize the second-order frequency-domain Volterra kernel. However, higher-order products were not considered in this work and the Volterra model was used in real-time domain for a very low-frequency application.

In this chapter, an efficient method is proposed to discriminate between different individual IMD products [5]. By adopting the proposed approach, these products are located at separate frequency locations, each different from the location of the other IMD products. In this way, each single IMD product can be characterized separately in a single measurement. The fundamental differences between the proposed approach and the one used in [84] are that 1) the proposed method considers nonlinearity orders of 3 and higher; 2) through its efficient frequency manipulation it can characterize memory effects and nonlinearity order levels; 3) because of its efficient frequency allocation of the tones, the tone spacings are not much different from each other; and 4) it is dedicated to modulated RF applications, although it is more general and can be used for other applications.

Due to the distinction between IMD products in this approach, several advantages are obtained, including the ability for in-band distortion measurements, memory effects characterization and model performance assessment. These three along with some other advantages are discussed in this chapter.

Section 4.2 explains the proposed *offset multisine* method. The properties of this excitation is discussed in Section 4.3. Some tools with respect to this method are introduced in Section 4.4 and are used in Section 4.5 to show some applications of this method in the characterization of intermodulation and memory effects. Finally a conclusion is drawn in Section 4.6.

4.2 Frequency-offset multisine waveform

4.2.1 Theory of multisine offsetting

Without lack of generality, a multisine with an odd number of tones M is considered. Choosing an odd number is just because the representation of tones used in this section is more compact in this way. Nevertheless, the same theory holds for an even number of tones. In a representative measurement for prior work, the tone frequencies f_m's are chosen to be

$$f_m = F_c + mF_d \quad , \quad m \in Z \ , \ -\frac{M-1}{2} \leq m \leq \frac{M-1}{2}, \tag{4.1}$$

where F_c is the center frequency of the multisine and F_d is the frequency spacing between two adjacent tones. In this chapter, we refer to this multisine as a *conventional multisine*. It is known [12] that the IMD products located around the fundamental frequency are the odd-order products and their frequencies are:

$$f_{IMD,2k+1} = F_c + \left(\sum_{i=1}^{k+1} m_i - \sum_{j=1}^{k} n_j \right) F_d$$

$$m_i, n_j \in Z \ , \ -\frac{M-1}{2} \leq m_i, n_j \leq \frac{M-1}{2}, \tag{4.2}$$

Equation (4.2) shows that different combinations of IMD products, even from different orders, can have the same frequency locations [12]. Therefore, with this conventional multisine we are not able to distinguish between IMD products in a straightforward way. Fig. 4.1a and 4.1b demonstrate that third-order intermodulation (IM3) products of different combinations can share the same frequency position for in-band and out-of-band IMD, respectively. The proposed approach is to offset the tone frequencies by different amounts, in such a way that each of the IMD products has its own unique frequency, as illustrated in Fig. 4.1c and 4.1d. The amount of offset is considered to be as small as possible so that the offset tone can be a good representative of the original tone in the conventional format. To achieve this goal, we shift the tone frequencies based on the frequency grid spacing F_g which is the frequency resolution of the simulation or measurement. Equation (4.1) is then modified to:

$$\hat{f_m} = F_c + mF_d + p_m F_g$$

$$m, p_m \in Z \ , \ -\frac{M-1}{2} \leq m \leq \frac{M-1}{2}. \tag{4.3}$$

in which the last term represents a small shift applied to each tone frequency. The hat sign above f_m indicates the shift in frequency, and Z is the set of all integer numbers. If we choose p_m values so that each $\hat{f_m}$ is unique with respect to other $\hat{f_m}$'s, we will achieve our target. Note that F_d is a multiple of the grid frequency F_g. Therefore, the tone frequencies in eq. (4.3) are still commensurate, although they are now commensurate with a much smaller base frequency F_g, as opposed to the base frequency F_d in the conventional multisine setting. Equation (4.3) states that each tone of the offset multisine is localized by coarse allocation $(F_c + mF_d)$ followed by a small refinement. A computer code is generated to find a set of p_m values as small as possible satisfying the uniqueness requirement.

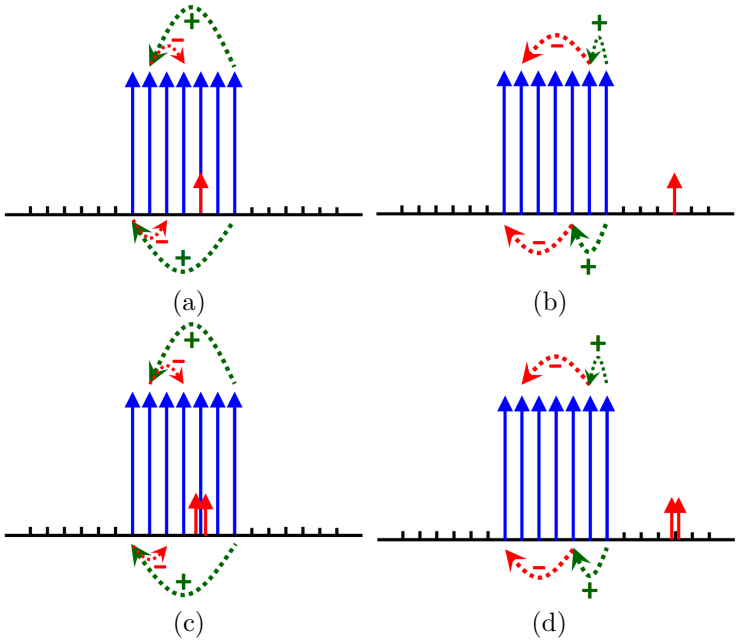


Figure 4.1: Formation of IMD products (vertical gray arrows): (a) In-band IMD products resulting from conventional multisine excitation share the same location as those of the main tones; (b) out-of-band IMD products from different combination of tones resulting from conventional multisine excitation share the same location; (c) in-band IMD products resulting from the offset multisine excitation; and (d) out-of-band IMD products resulting from the offset multisine excitation have distinct frequency locations. Here the dotted arrows connect frequencies that are combined to form 3^{rd} order products [5].

Due to the finer frequency grid of the offset multisine (F_g) as compared to that of the conventional multisine (F_d), the period of the offset multisine is much longer than that of the conventional one. However, this disadvantage can be insignificant, because the period can still be comparable with the length of a typical real modulated signal used in testing.

4.2.2 Simulation example

As illustration of the proposed method, we consider the following simulation example. A seven-tone multisine with around 2.4 MHz bandwidth is selected. The set of frequency locations of the tones with respect to the center frequency is $\{-1200, -812, -422, -28, 378, 798\}$ kHz, chosen so that each IMD product up to the fifth order has its own individual unique location. Since the aim of this section is to show the effect of offsetting the tones, the amplitude and phase manipulation of the tones are not considered here, and will be discussed in the subsequent section. The waveform is then used as an input to a nonlinear block with the characteristic function

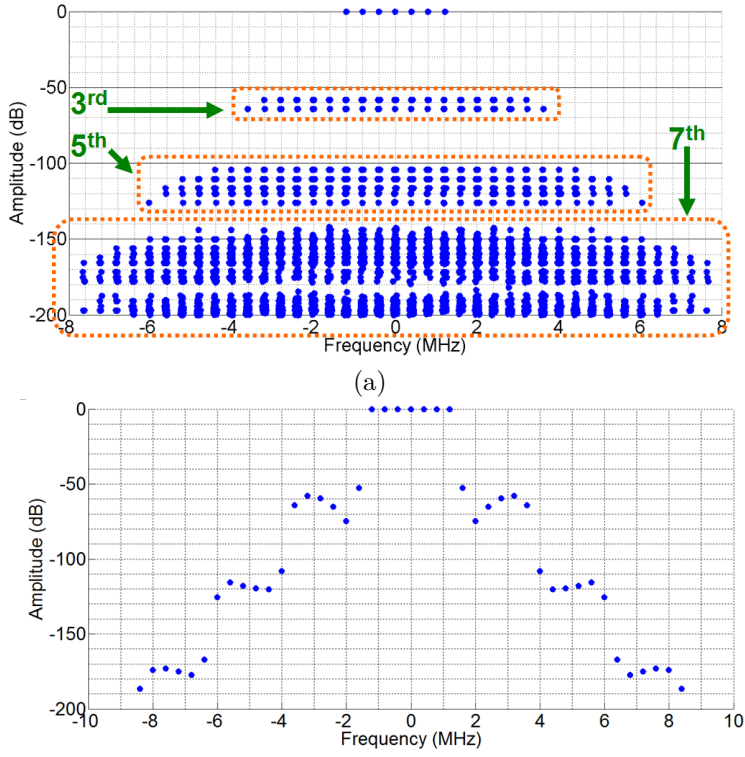
$$y = 4\left(1/\left(1 + e^{-x}\right) - 0.5\right). \tag{4.4}$$

The simulated output spectrum is plotted in Fig. 4.2a. Four distinct bunches of IMD products corresponding to the 3^{rd}, 5^{th}, 7^{th}, and 9^{th} order IMD products (denoted hereafter as IMD3, IMD5, IMD7 and IMD9, respectively) can be clearly seen. This is the particular feature of the proposed method that is absent in the conventional multisine approach whose respective spectrum is plotted in Fig. 4.2b. Note that the power level difference of different IMD orders changes with changing the signal power. Power levels of higher order IMD products increase faster than those of lower orders [12] as functions of the input power.

It can be seen in Fig. 4.2a that each bunch of IMD products has a different number of rows of frequency components. third-order intermodulation distortion (IMD3) has two rows, fifth-order intermodulation distortion (IMD5) has five rows and so forth. These numbers can be explained by examining the frequency-domain Volterra kernels [13] as follows.

IMD3 products correspond to terms of the type:

$$Y_3(\omega_1 + \omega_2 - \omega_3) = H_3(\omega_1, \omega_2, -\omega_3)X(\omega_1)X(\omega_2)X^*(\omega_3), \tag{4.5}$$



(a)

(b)

Figure 4.2: Output spectrum of eq. (4.4) with a) offset multisine excitation, and b) conventional multisine excitation. In (a) different nonlinearity orders of different degrees, even inside the main band, are easily distinguishable [5].

where H_3 is the frequency-domain Volterra Kernel. There are four cases of different angular frequency combinations: 1) $\omega_1 = \omega_2 = \omega_3$ 2) $\omega_1 = \omega_3$ or $\omega_2 = \omega_3$ 3) $\omega_1 \neq \omega_2$, $\omega_2 \neq \omega_3$, $\omega_1 \neq \omega_3$ and 4) $\omega_1 = \omega_2 \neq \omega_3$. Cases 1 and 2 lead to IMD products which are exactly at the same frequency as the main large tones, and are coherent with the tones [12]. Cases 3 and 4 generate products at distinct frequencies, and are incoherent with the tones [12]. However the products in case 3 can be created in two ways: $\omega_1 + \omega_2 - \omega_3$ and $\omega_2 + \omega_1 - \omega_3$. Therefore the products from case 3 are two times larger in voltage than the products from case 4. This is why we see two rows of IMD3 products with exactly 6 dB difference. The same reasoning applies to the existence of five rows of IMD5 products, as well as to the higher order products. In the IMD5 case, the levels of the five respective rows are weighted by 1, 2, 3, 6 and 12. Note that

in the case of the fifth order, we have two sets of products that lie at the same locations of lower order products. One set has the frequency combination of $(\omega_1, \omega_2, \omega_3, -\omega_2, -\omega_3)$, assuming that the ω values can be different or identical. The product of this combination lies at the location of the large main tones. The other set has the frequency combination of $(\omega_1, \omega_2, \omega_3, -\omega_3, -\omega_4)$ assuming that ω_4 is different from ω_1, ω_2 and ω_3. The product of this combination lies at the locations of the IMD3 products. These products that are located at the frequencies of the lower order terms normally are much smaller than them. We can estimate the values of these products by the most similar distinguishable IMD5 products. Since the number of these special products is low as compared to the rest of IMD products, this estimation will not affect the accuracy of the technique.

4.3 Properties of offset multisine excitation

4.3.1 Advantages

There are several advantages for the offset multisine excitation. Three of them are seen with regard to in-band distortion, IMD products, and waveform statistics, as explained in this section.

In real modulated, as well as conventional multisine waveforms, in-band distortion products are located at the same frequencies of the main signal. Therefore, they cannot be measured independently. However, in our proposed multisine excitation, due to the specific tailoring of the tone frequencies, the incoherent in-band distortion products are at distinct frequencies, as already shown in Fig. 4.1c and in the measured results of Sections 4.4 and 4.5.

The frequency of each incoherent IMD product is unique in the proposed excitation. This uniqueness can help in the characterization of phenomena in nonlinear circuits. For instance, one can easily assess the contribution of each nonlinearity order, as explained in Section 4.4.2. Furthermore, as will be discussed in detail in Section 4.5, memory effects can be well characterized.

The unique properties of the offset multisine, including its waveform statistics discussed in the following subsection, make it very appropriate for IM3 characterization of the circuit. The comparison between IM3 responses with offset and two-tone excitations is shown and discussed in detail in Section 4.5.2.

4.3.2 Waveform statistics

The offset multisine is in fact a multisine with a very fine frequency grid, but with most of the tones switched off. Therefore, it effectively behaves like a multisine with a high number of tones in some aspects, although it has a few tones. One of the aspects is the real-imaginary (RI) plot of the baseband signal. The RI plot is one of the tools to observe how populated the state space [13, 85, 86] of the waveform is. In systems including long-term memory effects, if a multisine is supposed to excite the same characteristic behavior of a nonlinear system as the real modulated waveform does, the former should have baseband state space coverage as close as possible to the latter.

The offset multisine has a highly populated RI plot, whereas a conventional multisine with the same number of tones has significantly less dense RI plot. Figures 4.3a and 4.3b show RI plots of seven-tone excitations, illustrating the large difference between the plots of the offset and conventional multisine cases with equal amplitude tones. In the conventional multisine case, the tone phases were manipulated for more population enrichment. The curve would be even simpler, i.e., a straight line, if the tone phases were equal. The fact that the conventional multisine has a simpler RI plot can be understood by considering that it has much smaller time period than the offset multisine due to its larger frequency grid. Therefore, it sweeps a smaller fraction of the plane in RI plot because it has a shorter time for variation.

The highly populated RI plot of the offset multisine suggests that it can be used as a suitable alternative for many real modulated waveforms, which also have populated RI plots. This means that the offset multisine is able to excite much more characteristic behaviors of the circuit, as compared to the conventional multisine with the same number of tones. In this sense, it is more similar to the real modulated waveform. Figure 4.3c shows an RI plot of a typical LTE waveform, with the same bandwidth as those of the multisine excitations in Fig. 4.3a and 4.3b. We can see that with even a small number of offset tones, e.g. 7 in this case, we can achieve a well-populated RI plot as compared to the plot of the real modulated waveform.

In addition to the qualitative inspection of the offset multisine behavior with RI plot, we can also examine some quantitative metrics.

After determining the tone frequencies, one usually determines the amplitudes and phases of the tones based on the PAPR or the complementary cumulative distribution function (CCDF) [48] criteria. Conventional multisine excitations are very convenient to manipulate in order to get the desired CCDF or PAPR.

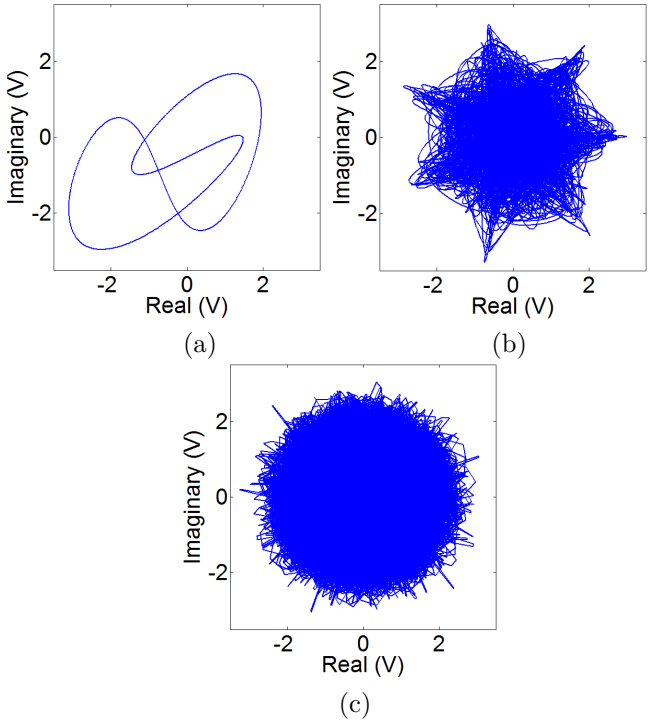
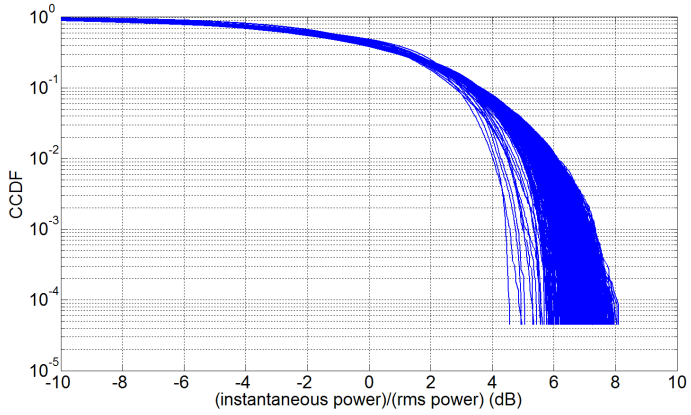


Figure 4.3: RI plot of a) conventional seven-tone waveform, b) offset seven-tone waveform, and c) LTE waveform.

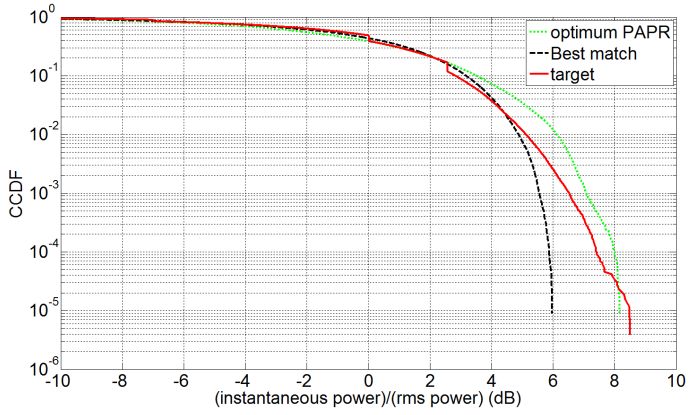
Now we show that offset multisine excitation, like the conventional multisine excitation, has enough degrees of freedom to get various CCDF or PAPR values. Therefore, because of this property and considering the abovementioned RI plot population of offset multisine excitations, we conclude that they have required statistical properties to be good representatives of real modulated waveforms.

Fig. 4.4a shows the CCDF curves of offset multisine with random selection of tone amplitude and phase. 500 realizations are made and plotted together in this plot. From this figure we can see that by varying amplitude and phase, we can change the PAPR by more than 3.5 dB.

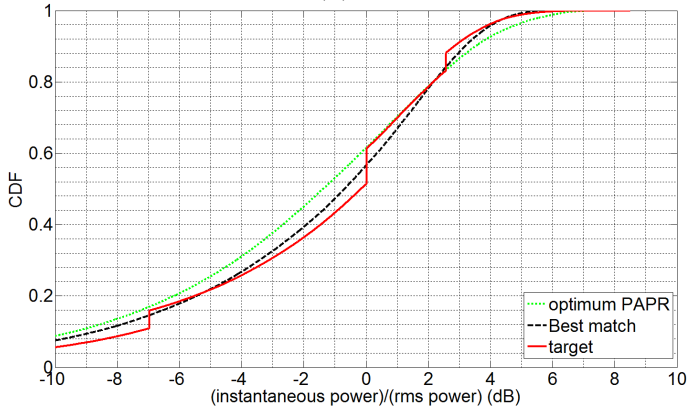
As an example of how we can determine the amplitudes and phases of the tones in the offset multisine, we chose an long-term evolution (LTE) waveform with 10 MHz bandwidth. A computer code is run to randomly change the amplitudes and phases of the tones and to select the optimum values. The



(a)



(b)



(c)

Figure 4.4: CCDF and CDF plots of offset multitone with different tone amplitudes and phases versus the instantaneous power normalized by the average (rms) power: a) whole ensemble curves b) optimum CCDF curves, and c) optimum CDF curves [5].

Kolmogorov-Smirnov test [87] is used to have a quantitative assessment about how similar the multisine distribution function is to that of the LTE signal. This test provides a criterion for the similarity of cumulative dstribution function (CDF) of two statistical samples.

Fig. 4.4b and 4.4c illustrates the CCDF and CDF plots of the LTE waveform, as well as the best candidate of the multisine ensemble in terms of CDF similarity to that of the LTE signal. This figure shows that the best candidate does not necessarily have the closest PAPR to the LTE signal.

In the subsequent sections, we use multisine with equal amplitudes. Note that if unequal-amplitude offset multisine is used, the IMD products should be normalized by their constitutive tones, in order to be able to use the proposed approach of this chapter.

4.4 Tools for characterization of IMD products

In this section we introduce the tools that we use for our investigation. The experimental setup is explained first. Then the plots used for characterization are presented with some experimental data. These plots are then used in the next section for the specific application of memory effects characterization.

4.4.1 Experimental setup

Measurement setup

The measurement setup used in this study is shown in Fig. 4.5. It mainly consists of a computer (PC), a vector signal generator (ESG), a PA, an attenuator and a vector signal analyzer (MXA). The waveform is loaded to the ESG as baseband IQ data. It is then converted into an RF waveform, and after passing through the PA, it is captured and downconverted to baseband IQ data in the MXA. The captured IQ data is then processed on the PC.

Three different PAs are used in our investigations. Two PAs are handset PAs for band 5 at 836 MHz (PA1) and band 13 at 782 MHz (PA2), both with maximum power of 27.5 dBm, and the third one is a base station PA module set around

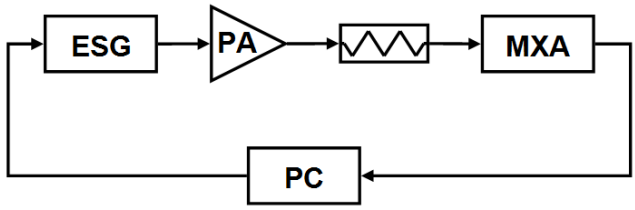


Figure 4.5: Measurement setup used in this chapter [5].

2.2 GHz(PA3) with the maximum peak power of 30 W. The PA gains are 28 dB, 32.5 dB and 15 dB respectively.

Waveform generation

Unlike conventional multisines, the offset multisine has unequal frequency spacing between tones. Therefore, finer grid size is required to generate such a waveform. Consequently, the period of an offset multisine is longer than that of a conventional multisine. For example, for a seven-tone offset multisine with a bandwidth of 2.4 MHz and a grid size of $F_g = 2$ kHz, the p_m values in eq. (4.3) can be:

$$(p_1, p_2, p_3, p_4, p_5, p_6) = (0, -6, -11, -14, -11, -1, 43) \tag{4.6}$$

These values are corresponding offsets from the frequencies ($-3F_d$, $-2F_d$, $-F_d$, $0F_d$, $+F_d$, $+2F_d$, $+3F_d$), respectively. Note that these p_m values are not unique and can have different variations; the set of values in eq. (4.6) is only one of the possible realizations. The smaller the p_m values, the more uniform the frequency spacings between the offset tones are. The number of tones is not restricted to seven and can be increased. However, as the number of tones increases, the search algorithm for finding the right values of p_m in eq. (4.3) gets harder exponentially. Fortunately, choosing the number of tones as large as seven or eight is enough to get PAPR values typical of standard modulated signals. It is desirable to have the same PAPR so that we can be confident that the circuit behavior is similar to that of the case where we excite it with a real modulated signal. In our case we chose seven tones which have a PAPR of about 8 dB.

As a convenient way for signal generation, we first generate our desired tone in the frequency domain, and then take an inverse Fourier transform (IFT) to get

the time domain signal. In this way we are sure that our generated signal is cyclic.

4.4.2 Plots of IMD vs. frequency

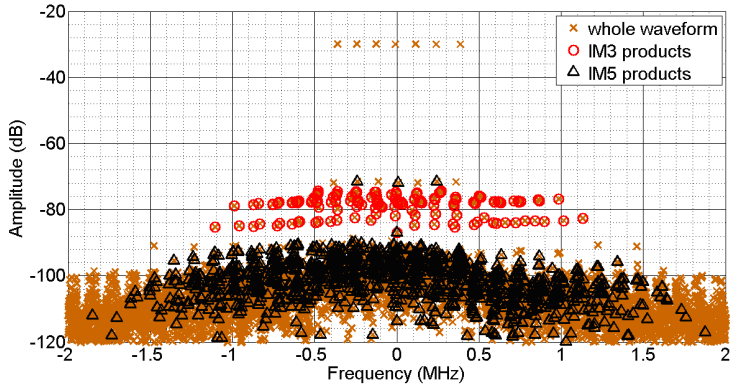
In this section, the PAs are excited with the offset multisine excitations with varying bandwidth. As described in this section, the captured data are analyzed based on two plots: 1) Output spectrum plot and 2) IM3 profile.

Spectrum plot

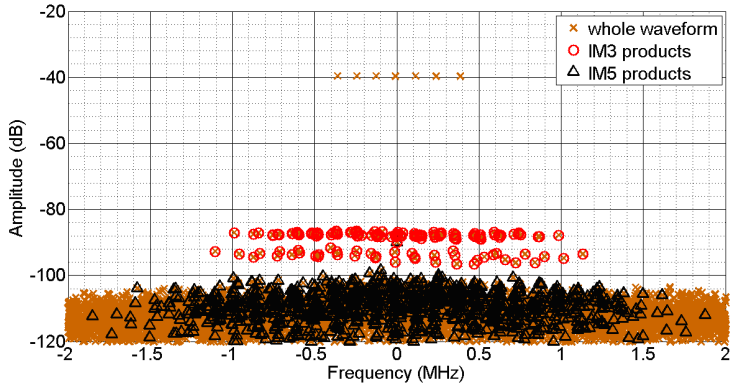
For each excitation with a specific bandwidth, the PA output spectrum is analyzed. Fig. 4.6 shows the spectra for the three PAs with the same excited bandwidth of 720 kHz. In these plots, the IMD3 and IMD5 products are marked with circles and triangles, respectively, whereas the entire set of data points, including IM3 and IM5, are all marked with crosses. The plots clearly show that the two rows of IM3 for these PAs are distorted as compared to those of Fig. 4.2a: They are not perfectly flat anymore. This phenomenon is due to memory effects, as will be discussed in the following. It is also clear that the amount of distortion fluctuation varies from one PA to another one, which is due to different memory behaviors of different PAs.

Since the IMD products are distinct from each other, by observing the output spectrum excited by the offset multisine, one can assess the amount of each nonlinearity order. Note that for IMD analysis in this chapter nonlinearity orders up to five are considered, like in Section 4.2.2, and the conditions of interest are such that the higher orders are negligible. However, the approach can also be used for higher order nonlinearity, provided that the offset multisine is modified to take into account the uniqueness of higher IMD locations.

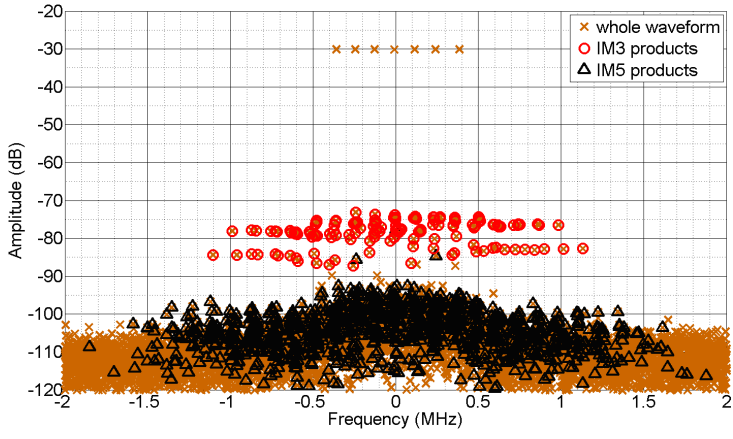
In Fig. 4.6a, the IMD3 products are much larger than the IMD5 products. Hence we can assume that those IMD5 products that share the same frequencies as IMD3 products, as explained in Section 4.2.2, do not have considerable impact on the IMD3 levels. This means that as long as the IMD3 products are substantially larger than the IMD5 and other unwanted products, by changing the PA output power, the IMD3 products pattern changes in level, not in shape. This fact is easily verified, but not shown here due to space limitation. However, if we increase the power, the IMD5 products start to affect the IMD3 levels.



(a)



(b)



(c)

Figure 4.6: Measured baseband output spectrum of a) PA1 at 21 dBm output power, b) PA2 at 11 dBm output power, and c) PA3 at 21 dBm output power. The PAs are excited with a seven-tone offset multisine with 720 kHz bandwidth. Different intermodulation products from different nonlinearity orders are distinguishable in the offset multisine setting [5].

The IMD5 products that share the same location as the lower IMD3 products are the results of $\omega_1+\omega_1+\omega_3-\omega_3-\omega_2$ and all its combinations, where $\omega_1 \neq \omega_2$. The number of possible combinations is 11. This number is found in a way very similar to what is presented in Section 4.2.2. Therefore, this specific accumulated IM5 level is at most 11 times higher than the level of the lowest IM5 row, which is itself composed of only one IM5 product. In other words, eleven IM5 products lie at the same frequency as that of each IMD3 product. Thus, in order that the IMD5 has impact on the IMD3 products, the sum of eleven IMD5 products should be comparable with the IMD3 level. The best way to estimate this level is using the level of the highest row of IM5, which is 12 times (21.6 dB) larger than the level of a single IMD5 product (corresponding to the lowest IM5 row). Based on these numbers, the accumulated strength of the IM5 products that are impacting each IM3 product is at most around $20\log(12/11) = 0.8$ dB lower than that of the highest IM5 row. So the highest IM5 row is a good measure to see how influential is IMD5 distortion on IMD3 products. All these discussions are under the assumption that the nonlinearity orders higher than five are negligible; otherwise we should extend our discussion to consider their effects as well.

Figure 4.7 illustrates the spectrum plots of PA1 output at 24 dBm power. The shape of the IM3 pattern is clearly distorted, as opposed to Fig. 4.6a, due to large IM5 products whose power levels are now comparable with those of IM3 products: The highest IM5 level has even exceeded the lower IM3 level. Furthermore, the IM3 level is not increased by $3 \times (24 - 21) = 9$ dB, again due to the IM5 impact. Note that the IMD3 and IMD5 products usually add up destructively.

Hereafter in this chapter, we focus on the cases that IMD3 products are sufficiently larger than other unwanted products, including IMD5. In this way, we can assume that, to a good approximation, the observed IM3 levels have resulted only from third order nonlinearity, and not higher.

In both Fig. 4.6a and Fig. 4.7, a row of seven products closely above the higher IM3 row can be seen. These are not the IMD products, but the image of the multisine due to the imperfect image suppression of the IQ-modulator inside the ESG. These products can be suppressed by balancing the I and Q signals in ESG through the calibration process. This calibration is done for Fig. 4.6b and 4.6c, and no images can be seen anymore in these plots. However, we deliberately include the imbalance IQ case in Fig 4.6a and 4.7 to show another advantage of the offset multisine: If there is any image lying inside the band, they have frequency locations distinct from the main tones and its effect cannot be mistaken for other factors, such as modeling error in modeling application.

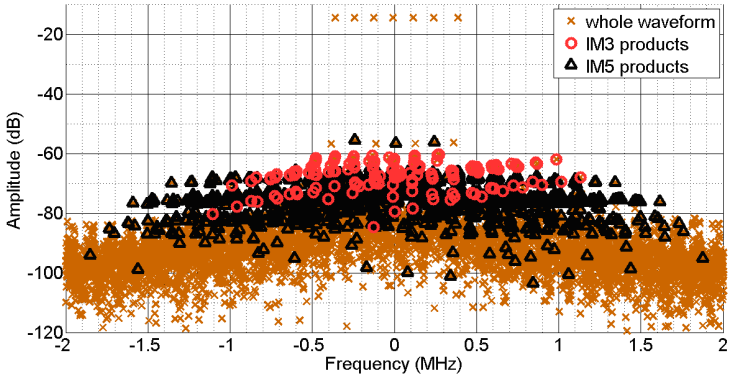


Figure 4.7: Measured baseband output spectrum of PA1 at the output power of 24 dBm. The PA is excited with a seven-tone offset multisine with 720 kHz bandwidth [5].

Figure 4.8 shows the IM3 spectrum of PA3 at 23 dBm output power with 1.2 MHz bandwidth. IMD3 products are marked based on $\omega_2 - \omega_3$ in eq. (5). In this figure, Δm is the difference between the coefficients m of $\hat{f}_2$ and $\hat{f}_3$ in eq. (4.3), where $\omega_2 = 2\pi\hat{f}_2$ and $\omega_3 = 2\pi\hat{f}_3$. As explained in Section 4.2.2, there are basically two rows of IMD3 products. Each point in the upper one is the summation result of two combinations $\omega_1 + \omega_2 - \omega_3$ and $\omega_2 + \omega_1 - \omega_3$. That is why each point in the upper row has more than one marker at its location. However, the lower row consists of the points purely resulting from $\omega_1 + \omega_1 - \omega_2$ and no other combination. Henceforth, we call this row the 'lower IM3'. The points from the lower IM3 are circled in Fig. 4.8. This purity makes the points of the lower IM3 suitable for our investigation. In fact, this row bears the information of the dependence of IM3 on the frequency difference $\omega_1 - \omega_2$: Considering the markers, the row shows that the IM3 level depends on this difference. To make this phenomenon clearer, we plot the lower IM3 profile in a different way in the following section.

IM3 profile

In the next step, we sweep the multisine excitation bandwidth, gather IM3 data from all these measurements, and plot the lower IM3 data points versus frequency difference. The resulting plot is shown in Fig. 4.9. In this figure, the horizontal axis is the frequency difference $(\omega_1 - \omega_2)/2\pi$, which can also be negative. This is another way of demonstrating IM3 curves for memory effects

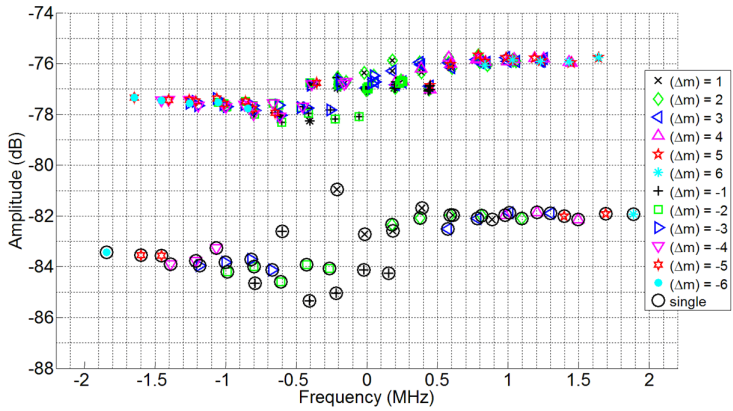


Figure 4.8: Measured IM3 spectrum of PA3 at the output power of 23 dBm, excited with a seven-tone offset multisine with 1.2 MHz bandwidth.. The IMD3 products are marked separately based on their Δm values, where Δm is the difference $m_2 - m_3$. The coefficients m_2 and m_3 are the respective values of m in (3) for f_2 and f_3, from which the IM3 frequency is found by $f_{IM3} = f_1 + f_2 - f_3$ [5].

as opposed to the typical one, in which the upper and lower IM3 curves are plotted on top of each other versus the absolute value of frequency difference.

Figure 4.9 clearly shows both asymmetry and unevenness of the IM3 profile, which are the consequences of slow (long-term) memory effects.

It is well known that, assuming negligible nonlinearity order higher than three, the transfer function of the IM3 product at $2\omega_2 - \omega_1$ is the summation products of three terms [88, 89, 90]. That is:

$$H_{IM3}(\omega_2, \omega_2, -\omega_1) = H_3(2\omega_2 - \omega_1) + H_2(2\omega_2) \cdot H_1(-\omega_1) + H_2(\omega_2 - \omega_1) \cdot H_1(\omega_2) \tag{4.7}$$

where H_{IM3} is the IM3 transfer function, and H_1, H_2 and H_3 are the first, second and third order transfer functions, respectively. Correspondingly, the IMD3 product at $2\omega_1 - \omega_2$ is:

$$H_{IM3}(\omega_1, \omega_1, -\omega_2) = H_3(2\omega_1 - \omega_2) + H_2(2\omega_1) \cdot H_1(-\omega_2) + H_2(\omega_1 - \omega_2) \cdot H_1(\omega_1) \tag{4.8}$$

If the tone spacing between ω_2 and ω_1 is small compared to the system bandwidth, we can assume that $H_3(2\omega_2 - \omega_1) \approx H_3(2\omega_1 - \omega_2)$, $H_2(2\omega_2) \approx$

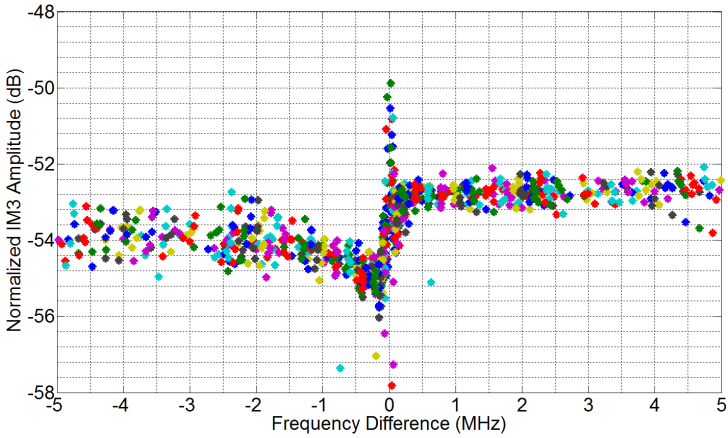


Figure 4.9: Measured IM3 profile of PA3 output at the output power of 23 dBm. Different colors (or color densities in grayscale version) in the plot indicate data points collected from different measurements [5].

$H_2(2\omega_1)$ and $H_1(\omega_1) \approx H_1(\omega 2)$. However,

$$H_2(\omega_2 - \omega_1) = H_2^*(\omega_1 - \omega_2) \neq H_2(\omega_1 - \omega_2), \tag{4.9}$$

where the asterisk denotes conjugation operation. These equations theoretically show that IM3 profile asymmetry happens if the imaginary part of the third term, as well as the imaginary part of the rest of eq. (4.7) and eq. (4.8), are not negligible [91].

Based on eq. (4.7) and eq. (4.8), the unevenness of the IM3 profile also happens when $H_2(\omega)$ around DC is frequency dependent and is large compared to $H_2(2\omega)$, which is a likely situation due to bias circuits of PAs [89, 90]. Note that second-order products around DC can modulate the PA bias due to the low-frequency response of the bias circuit.

Some spread can be seen in Fig. 4.9, which is due to measurement accuracy limitation, as well as a small IM5 effect. These could be reduced in future work. However, the global shape is clear enough for the assessment of slow memory effects.

Note that with a real modulated waveform that has a continuous frequency spectrum, we are not able to distinguish the dependence of IM3 on frequency difference. In the continuous spectrum with the bandwidth (BW) we have every frequency difference between two components of the spectrum varying from

zero to BW, and the IMD products share the locations. Therefore, we cannot extract the IM3 profile with a continuous-spectrum signal. This drawback highlights one of the advantages of the offset multisine: While the statistics of the offset multisine are close to those of the real modulated waveform, the offset multisine provides some useful insight into IMD behavior of the circuit which is not provided by real modulated waveforms. This advantage will be reconfirmed in Section 4.5.3, where the offset multisine provides better assessment of model strength to predict IMD behavior.

4.5 Focused application: memory effects characterization

As an interesting application of the proposed excitation, we focus on memory effects characterization. However, the approach is much more general and can be applied to different nonlinear characterization applications.

4.5.1 Long-term memory characterization

In the following, the methodology is applied to long-term memory effects investigation. The method can also be easily applied to short-term memory effects investigation, provided that we have a sufficiently wideband data-capturing equipment relative to the signal bandwidth.

We already showed through eq. (4.7)-eq. (4.9) in Section 4.4.2 that the frequency dependence of $H_2(\cdot)$ around DC results in unevenness and asymmetry of the IM3 profile. One of the major factors determining $H_2(\cdot)$ around DC is the bias circuit. This section investigates the effect of this factor on long-term memory and the IM3 profile.

Figure 4.10 shows the schematic of the GaN amplifier PA3 module mentioned above, with operation near 2 GHz, including its bias circuits. We show that we can efficiently inspect memory effects with the offset multisine technique, including changes induced by intentional manipulations of the bias networks. Figure 4.9 is the IM3 profile of the complete PA3 circuit. Figure 4.11 shows the IM3 profiles after various intentional changes to the bias network. Different colors (or color densities in grayscale version) in the plots indicate data points collected from different measurements. In Fig. 4.11a, the capacitor C15 is removed. This profile shows that C15 has a crucial roll in low-frequency response of the

circuit around DC. Due to nonlinearity of the transistor, the high-frequency components are mixed together and downconverted to frequencies around DC ($H_2(\omega_1 - \omega_2)$). If these downconverted products are not well suppressed in the bias line, they can modulate the DC supply of the transistor, be mixed with the RF large tones and upconverted back to the high frequencies around the main band (reflected in $H_2(\omega 1 - \omega_2)H_1(\omega 1)$). C15 suppresses the downconverted product by bypassing it to the ground. By removing C15, we allow this mixing mechanism to have its effect on the low-frequency IM3 profile, in an M-shape variation visible in Fig. 11b. In other words, C15 reduces the amplitude of $H_2(\omega_1 - \omega_2)$. If C15 is removed or not large enough, $H_2(\omega_1 - \omega_2)$ is considerable and its frequency dependency highly affects the IM3 profile, leading to long-term memory.

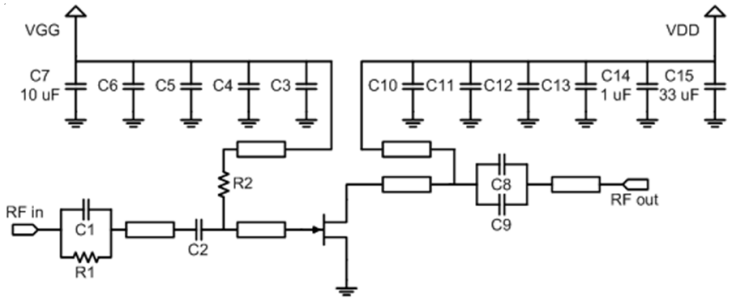


Figure 4.10: Simplified schematic of PA3 module (CREE CGH27030F) [5].

To further confirm that the IM3 profile shape is due to the bias line circuitry, we replace C15 with an RLC circuit. The resonant frequency of the RLC is around the peak frequencies of Fig. 4.11b, with the bandwidth of around 60 kHz. Since C15 is removed, without the RLC, the IM3 profile should be like Fig. 4.11b. However, since the RLC suppresses the signal around its resonance frequency, the profile of Fig. 4.11c is obtained.

Figure 4.11d is the IM3 profile after removing both C15 and C14. Comparing this figure with Fig. 4.9 and Fig. 4.11a demonstrates that C14 suppresses the peak observable in Fig. 11d. Figures 4.11e and 4.11f are the IM3 profiles with C7 = 100 nF and removed C7, respectively. These two figures highlight the importance of C7, a capacitor in the gate bias line, in memory effects and IM3 rejection.

In summary, the plots of Fig. 4.11 provide information about the impact of bias line components on IM3 level variation and long-term memory effects. There

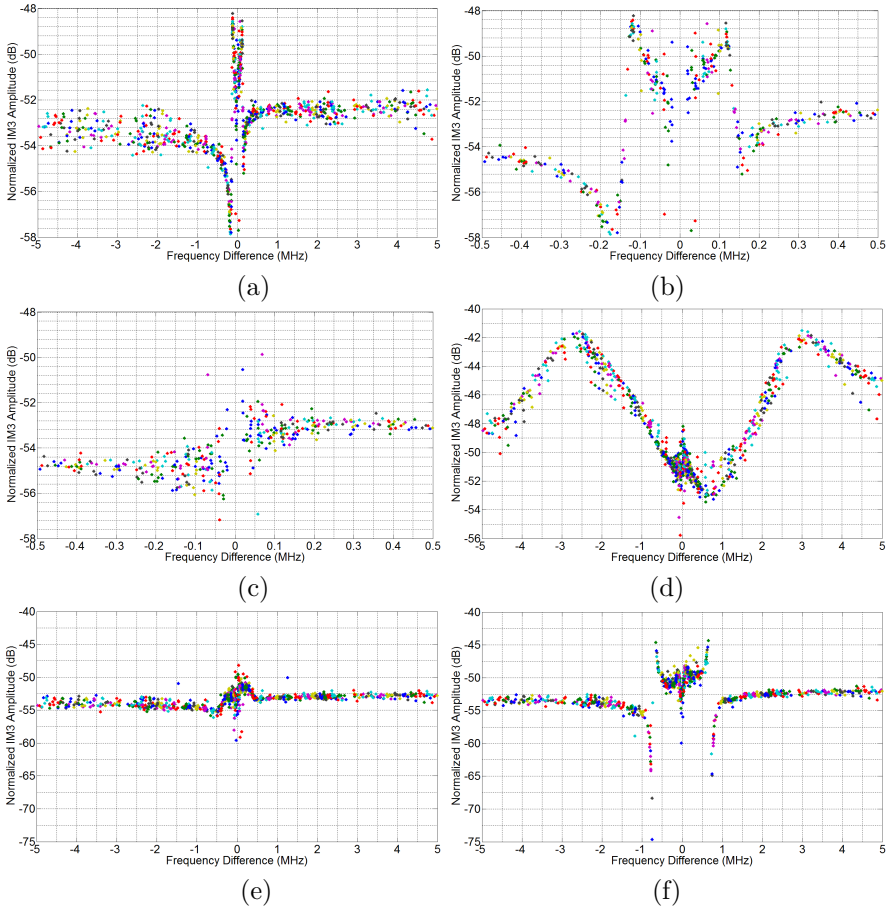


Figure 4.11: Measured IM3 profile of PA3 module with a) C15 removed, b) C15 removed (zoomed view), c) C15 replaced with RLC, d) both C15 and C14 removed, e) C7 replaced with a 100 nF capacitor, and f) C7 removed. Different colors (or color densities in grayscale version) in the plots indicate data points collected from different measurements [5].

are articles in the literature, for example [91, 92, 93, 94, 95], explaining this phenomenon in more detail.

The proposed offset multisine is much more suitable than the conventional multisine or modulated signals for investigation of long-term memory effects due to two main reasons: 1) in wideband applications, long-term memory effects show themselves in in-band frequency components and in others close to the main band. This localization is because of the fact that for long-term memory effects, the subtraction term in $\omega_1 + (\omega_2 - \omega_3)$ has a small value as compared to the tone frequencies. Because of this localization, we must use a method to measure in-band distortion in order to be able to properly characterize long-term memory effects; 2) in the conventional multisine or real modulated waveform, we can see only the accumulated effect of all IM3 products at each frequency, and we cannot separately study the individual products that lie at the same frequency. However, in our method, this investigation is straightforward.

4.5.2 Comparison with two-tone excitation

Two-tone excitations have been widely used to characterize the IM3 response of nonlinear circuits. This section discusses the two-tone applicability and highlights another advantage of our proposed excitation: The two-tone excitation is not very appropriate for IM3 characterization, in the sense that the waveform statistics do not have enough similarity to real modulated waveforms. However, the offset multisine is more appropriate for IM3 characterization, since it has more similar statistics to real modulated signals.

For amplifier PA3, when C7 was removed, another measurement was conducted using two-tone excitation, for comparison. The bandwidth was swept and all the IM3 data were collected to create an IM3 profile for the two-tone excitation. The profile is shown in Fig. 4.12. It is evident that two-tone excitation also shows variation in IM3 profile, but the peak locations are not the same as those of Fig. 4.11f. Therefore, since our offset multisine has statistics more similar to the real modulated signal than the two-tone waveform, we can conclude that although the latter is able to characterize the IM3 behavior to some extent, its characterization is not as representative as that of the offset multisine.

In order to make the two-tone excitation more similar to a real signal, we have to increase the number of tones, which typically leads to the conventional multisine. However, conventional multisines are improper in the sense that IM3 products share the same frequency with each other or with the large tones. This sharing hinders appropriate memory effect characterization, especially for very

long-term memory characterization which requires information about in-band distortion. Note that for long term memory effects $\omega_2 - \omega_3$ is small. Therefore the resulting product, $\omega_2 + \omega_2 - \omega_3$, lies near main tones.

Martins et al. [89] demonstrated that two-tone characterization can be used to predict the multisine response. They experimentally verify this hypothesis with a PA over a certain bandwidth. However, they did not investigate what happens if they sweep the bandwidth in a wide range. Comparison of Fig. 4.11f and Fig. 4.12 suggests that two-tone characterization can be properly used for predicting offset multisine behavior, providing that the tone spacing is not around the peak point or at the location of large variation. If the tone spacing is near these frequencies, the two-tone test cannot predict the baseband response of the circuit well.

Another advantage of the offset multisine over the two-tone excitation is that multiple data points in the IM3 profile are obtained by each measurement. For instance, more than 40 points in the IM3 profile are obtained in the frequency difference range of F_d and $7F_d$ (see eq. (4.3)), if a seven-tone multisine excitation is used.

Another superiority of the offset multisine over the two-tone excitation is that we can measure different IM5 products separately, whereas in the two-tone case many IM5 products share the same frequency location because there are only two distinct tones in this case. Therefore, the two-tone excitation is not as appropriate as the offset multisine excitation for IM5 characterization.

4.5.3 Figure of merit for model performance

Since the proposed excitation gives more details about intermodulation products than the conventional multisine does, it can be used as a suitable figure of merit to assess the performance of nonlinear models. By using this excitation as our test waveform, one can examine the detail of the memory effect and nonlinearity orders of the system and evaluate if they are predicted properly by the model.

As an example we examined four different models with the offset multisine excitation: memoryless, memory polynomial (MP) [61], 1st order dynamic deviation reduction (DDR) [96] and piecewise dynamic deviation reduction (PW-DDR) [97] models. The parameters of the models are chosen based on model improvement. For each model, the number of coefficients is increased until the amount of improvement is negligible, namely less than 0.1 dB in our case. Then the results are compared together. The selected parameters are

Table 4.1: Comparison of normalized root mean square error between predicted and measured waveforms

Model	Memoryless	MP	DDR	PW-DDR
NRMSE (dB)	-38.4	-40.4	-40.9	-41.0

$P = 9$ and $M = 0$ for the memoryless model, $P = 9$ and $M = 3$ for the MP, and $P = 13$ and $M = 3$ for the DDR. Here, P is nonlinearity order and M is memory depth in terms of sample time. For the PW-DDR we have chosen the magnitude thresholds as $\{0.1200, 0.5500\}$, the corresponding nonlinearity orders as $\{3, 5, 7\}$ and the memory depths as $\{2, 2, 2\}$ (see [97] for more explanation).

Figure 4.13 shows the measured and estimated results of these four models, respectively. Starting with Fig. 4.13a, the memoryless model predicts a flat frequency response for the single IM3 entries at around -45 dB and -51 dB levels. Fig. 4.13b shows the memory polynomial prediction, illustrating that the model estimates the general shape of the IMD rows, but it cannot estimate their detailed variation. This mismatch is best seen in the upper IM3 row. Both the DDR and PW-DDR have better modeling performance for the variation of IMD points. This superior performance can be seen in Fig 4.13c and 4.13d, which exhibit a better match between the IMD estimates and the measured data.

Table 4.1 shows the performance of the models in terms of NRMSE between the predicted and measured waveforms. As can be seen, the NRMSE values in Table 4.1 do not reflect the clear difference in IMD estimation capabilities of the models apparent in Fig. 4.13. In fact, NRMSE is based on a normalized average power error of the whole signal including the large main tones, whereas the IMD differences observed in Fig. 4.13 are at a much smaller signal level relative to the dominant seven-tone multisine excitation.

Due to this weakness of the NRMSE criterion to reflect the IM3 similarity between measurement and simulation, we define another figure of merit based on only IM3 spectrum. That is:

$$NRMSE_{IM3} = \frac{\sqrt{\sum |IM3_{est} - IM3_{meas}|^2}}{\sqrt{\sum |IM3_{meas}|^2}}, \tag{4.10}$$

where $IM3_{est}$ and $IM3_{meas}$ are the estimated and measured IM3 phasors of

Table 4.2: Comparison of IM3 normalized root mean square error between predicted and measured waveforms

Model	Memoryless	MP	DDR	PW-DDR
$NRMSE_{IM3}$ (dB)	-13.8	-17.0	-21.2	-22.2

the voltage spectrum, respectively. The calculated $NRMSE_{IM3}$ values are shown in Table 4.2.

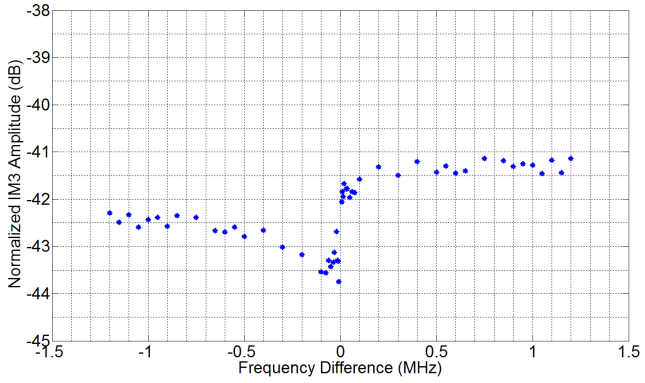
With this proposed figure of merit, now we can clearly see the performance differences of various models for IM3 products, which are consistent with the predictability observed in Fig. 4.13. As can be seen in the plots, PW-DDR has the best performance, and the corresponding $NRMSE_{IM3}$ value confirms this superiority. DDR has a performance better than MP, but worse than PW-DDR. The memoryless model has the weakest performance among the four models. These results can be justified by the fact that the deviation of the IM3 rows from flatness is due to memory effects and the model which has better capability to model memory effects gives better estimation.

These results show that, as mentioned in the end of Section 4.4.2, the offset multisine can distinctly show the predictability of different IMD products by models, while this assessment is not possible with the conventional multisine or real modulated excitation which has continuous spectrum.

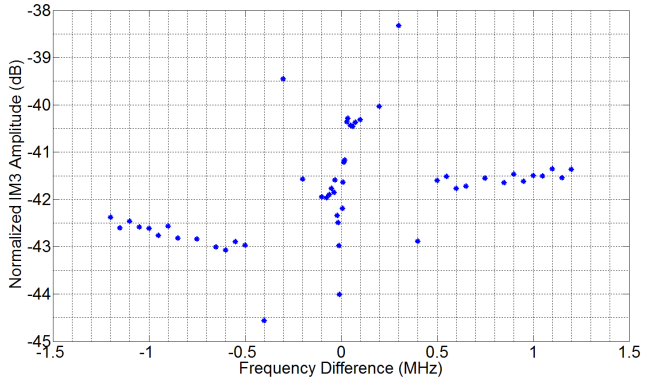
4.6 Conclusion

In this chapter we showed that by offsetting the tone frequencies of a conventional multisine in an appropriate way, we can have an excitation with a variety of useful properties. Its statistics can be more similar to those of a real modulated waveform. Moreover, due to the special choice of the tone frequencies, different IMD products have their own unique locations such that they can be distinguished from each other. This uniqueness enables interesting applications. It provides direct in-band distortion measurements, and separate measurements of IMD3, IMD5 and even higher-order IMD products. Therefore, it can show the strength level of each nonlinearity order. It also provides a strong memory effects characterization method which is faster and more reliable than classic two-tone experiments. Furthermore, opens a new horizon to the model performance

assessment, by assessing the model strength to estimate the memory effects. The effectiveness and usefulness of the approach have been successfully verified and demonstrated by a series of measurements. Another variation of offset multisine excitation will be presented in Chapter 6 to enable detection of influential nonlinearity orders in receive-band noise of power amplifiers.



(a)



(b)

Figure 4.12: Measured IM3 profile of PA3 at the output power of 23 dBm with two-tone excitation and a) complete bias circuit, and b) C7 removed [5].

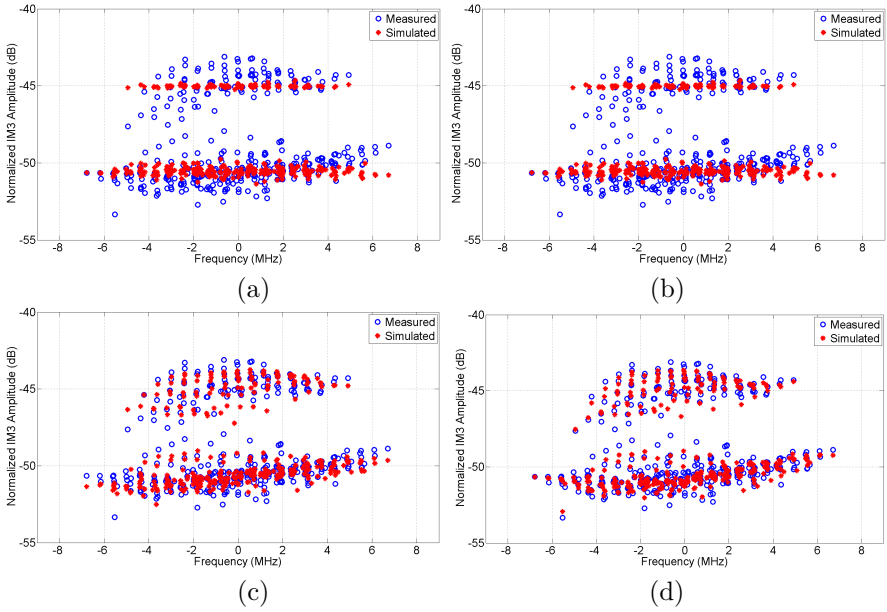


Figure 4.13: Measured IM3 spectrum of PA3 output and its estimation spectra with a) memoryless model, b) memory polynomial model, c) DDR model, and c) piecewise DDR model [5].

Chapter 5

Linearization of Data Converters

The contents of this chapter is based on the author's paper published in the IEEE Transactions on Microwave Theory and Techniques, November 2013 [6].

5.1 Introduction

Data conversion from the analog domain to the digital domain and vice versa is a critical process required in almost all telecommunication systems, precision instruments and modern electronic devices. One of the key issues in this conversion process is its accuracy. Due to the quantization process, some intrinsic quantization error always occurs. This error is also referred to as quantization noise. Quantization noise along with some other non-ideal effects in data converters are usually desired to be reduced. This chapter discusses some of non-ideal effects in data converters, and proposes a technique, called dithering, to mitigate such effects. As will be explained, the technique is applied to signal generation and is accomplished in post-processing.

Section 5.2 introduces dithering technique used in data converters for accuracy enhancement. Section 5.3 explains the basics of the proposed technique, namely ensemble averaging technique, in the static case. Then, the novel technique is presented in Section 5.4 to implement the ensemble averaging technique in the

RF characterization applications, supported by some simulation investigations. Some experimental investigations are then discussed in Section 5.5 for real-world evaluation of the technique. Finally conclusions are drawn in Section 5.6.

5.2 Dithering for accuracy enhancement

Generally, there are three methods to reduce quantization noise spectral power: (a) increasing the number of quantization levels by increasing the number of bits [98], (b) increasing the sampling rate [99], and (c) dithering [100, 101, 102, 103], which is the subject of this chapter. Dithering is the intentional addition of a waveform to the input signal prior to the data conversion. By a smart addition of dither, the resulting quantization noise of the total waveform can be decorrelated from the input signal. This decorrelation makes the quantization noise similar to white noise, and spreads the power of the quantization noise over the entire frequency range. This spread results in the reduction of the peak levels of the noise spectrum, at the cost of an increased noise floor. Figure 5.1 illustrates how dithering affects the quantization error and the quantizer output spectrum. First a simple sine wave x shown in Fig. 5.1a excites the quantizer. As can be seen in this figure, the quantization noise $q(x)$ is also periodic and has some correlation with the input x. That is why the quantizer output spectrum shown in Fig. 5.1c has some spurs at frequencies other than the fundamental frequency, namely ± 5 Hz. If enough noise is added to the input, as shown in Fig. 5.1b, the quantization noise will no longer be correlated with the input, and is no longer periodic. Hence the quantizer output spectrum, illustrated in Fig. 5.1d does not have the spurs anymore. However, the noise floor level is increased in this figure, as a result of the added noise.

Dithering techniques can be divided into two classes as illustrated in Fig. 5.2: subtractive and non-subtractive dithering [100, 101]. In subtractive dithering (Fig. 5.2a), the dithering signal $d(t)$ added to the input $x(t)$ is subtracted from the output signal, whereas in non-subtractive dithering (Fig. 5.2b) the dithering signal is not subtracted. Instead, because its frequency components are usually outside the frequency band of the input, they can be effectively filtered out [101]. Non-subtractive dithering is often preferred due to its simpler hardware.

In non-subtractive dithering, the dithering signal is usually chosen with a frequency that is distinguishable from the input signal's spectrum. For instance, the signal can be a sinusoid or band-limited noise [102]. However, white noise can also be used, followed by proper filtering after data conversion. In these methods, dithering is performed in the time domain, which means a time-varying

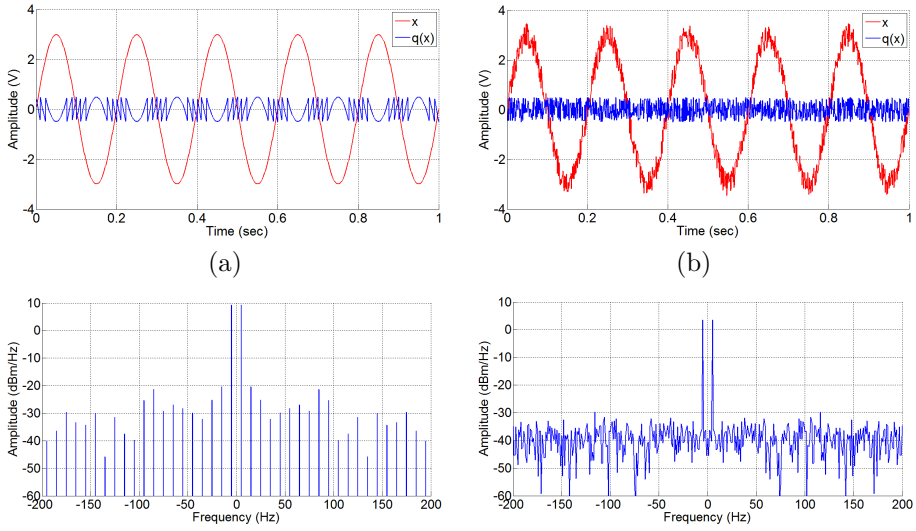


Figure 5.1: Demonstration of dithering effect on quantization noise and spectrum: The first row is the time plots of a sine wave x a) without, and b) with noise dithering along with its quantization noise $q(x)$. The second row is the spectrum plots of the quantizer output in the two cases of (a) and (b).

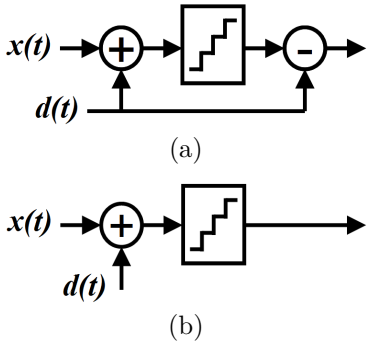


Figure 5.2: Dithering classes: a) subtractive, and b) nonsubtractive dithering. The blocks with the staircase represent quantizers.

dithering signal is added to the time-varying input signal. However, dithering can also be applied over an ensemble of identical signals [103]. In this approach, different realizations of the dithering signal are added to an ensemble of identical input waveforms. The whole ensemble is then averaged in post-processing to obtain a waveform with reduced quantization noise. A simple demonstration of this approach is shown in Fig. 5.3. In this figure, three identical sine waveforms are added with three independent noise waveforms. After quantization, the three quantized signals can be different from each other due to the added noise dithering signals that cause them lie in different quantization levels. After averaging, the resulting signal can have an average value that is more similar to the original input value as compared to the quantized levels, because the average of an ensemble of noise signals tends to zero. In practice, ensemble dithering requires acquisition of different realizations, which can be a time-consuming process.

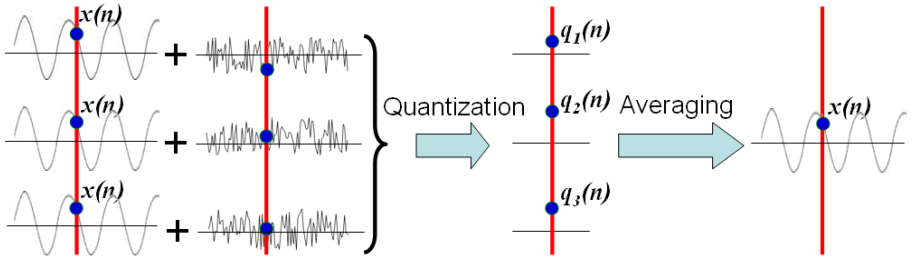


Figure 5.3: Ensemble averaging of signals dithered with noise.

In this chapter, we propose a time-domain ensemble-averaged dithering technique to reduce quantization noise. A dithering signal, which has a frequency much lower than a desired periodic waveform, is added over several periods of the waveform. The waveform is then segmented and the ensemble average of the segments is computed in post-processing. In this way, only one run of the measurement is required. Moreover, we use a simple deterministic sine function for dithering. The use of a sine function is shown to be an effective approach. This simple dithering technique can be used for test and characterization applications where high measurement accuracy—beyond what is introduced by the data converters in the measurement setup—is required. We show that by the use of large dithering signals in the proposed method we can compensate for some non-idealities of an arbitrary waveform generator (AWG). The outcome of this dithering approach is a post-processed signal with reduced impurity in its spectrum. Without dithering, the spectrum has impurity in the form of unwanted spurs spread over the whole spectrum [104]. If this impure signal is

used as an excitation for a nonlinear system, the spectral regrowth of the output may be masked by the spurs and not be diagnosed. However, the regrowth can be seen in the post-processed dithered signal, thanks to the reduced spurs.

5.3 Dithering over an ensemble

5.3.1 Quantization error and dithering

Without loss of generality for quantizers, we assume a mid-riser quantizer [105] with input x and quantizer step size Δ, as shown in Fig. 5.4. The output is then

$$Q(x) = \Delta[\frac{x}{\Delta}] + \frac{\Delta}{2} \quad , \quad X_{min} \leq x \leq X_{max}, \tag{5.1}$$

where $[\cdot]$ is the floor operator returning the greatest integer smaller than or

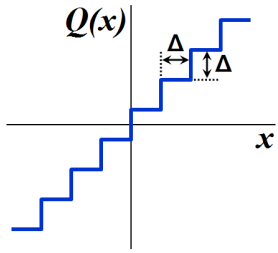


Figure 5.4: Mid-riser quantizer with step size Δ.

equal to its argument, and X_{min} and X_{max} are the lower and upper bounds of the quantizer, respectively. The quantization error $q(x)$ is then

$$q(x) = Q(x) - x, \tag{5.2}$$

which is periodic with the period of Δ over the valid range of the input. Without loss of generality, we can focus our investigations on the range $-\Delta/2 \leq x < \Delta/2$. We define a new variable u as the remainder of x after removing its integer *floor* part, i.e.,

$$u = x - [x]. \tag{5.3}$$

Assuming that $\Delta = 1$, eq. (5.1) and eq. (5.2) are rewritten as

$$Q(x) = [x] + \frac{1}{2} \quad \text{and} \tag{5.4}$$

$$q(x) = Q(x) - (u + [x]) = \frac{1}{2} - u \quad , \quad 0 \leq u < 1. \tag{5.5}$$

Since $0 \leq u < 1$, we find that $[u] = 0$. Thus,

$$q(x) = Q(x) - u = \frac{1}{2} - u = q(x). \tag{5.6}$$

Equations (5.5) and (5.6) show that for *static* analysis of a quantizer, we need to consider u only. The results will apply to the real signal x, because $q(x)$ is periodic with the period Δ. This is the approach that we employ here. The word *static* is used here to emphasize that we are not considering the quantization noise variation of the quantizer over time. Instead, we aim to reduce the quantization noise for different input states.

The quantization error $q(u)$ is plotted in Fig. 5.5 as a gray dotted line. Its range is $[-0.5, 0.5)$. Now, we show that by adding a suitable dither to an ensemble of waveforms and averaging them, we are able to reduce the quantization error. Note that the aim here is to reduce the static quantization noise, illustrated in Fig. 5.5. Therefore, an ensemble of waveforms is taken for analysis. Two different dithering functions are investigated: ramp and sine. We define these dithering functions as functions of ensemble sample number n, not functions of time. They are defined as

$$\text{ramp}(n) = \frac{n}{N_E} \quad , \quad n \in N \ , \ 1 \leq n \leq N_E \quad \text{and} \tag{5.7}$$

$$\text{sine}(n) = A_m \sin(2\pi n/N_E + \Phi) \quad , \quad n \in N \ , \ 1 \leq n \leq N_E, \tag{5.8}$$

where N_E is the ensemble size and A_m and Φ are the amplitude and phase of the sine, respectively. The dithered waveform is defined as

$$u_d(n, k) = u(k) + d(n), \tag{5.9}$$

where k is the time index and $d(n)$ is the dithering function. If we take the average of $u_d(n, k)$ over the ensemble, the result will be

$$u_{av}(k) = \frac{1}{N_E} \sum_{n=1}^{N_E} u_d(n, k) = u(k) + \frac{1}{N_E} \sum_{n=1}^{N_E} d(n). \tag{5.10}$$

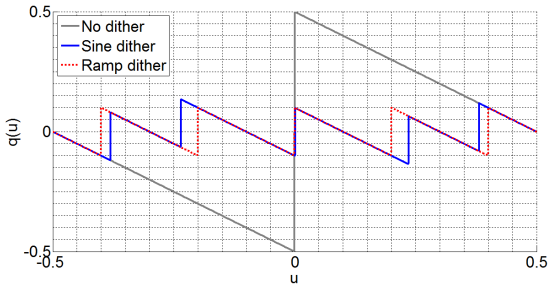


Figure 5.5: Quantization noise with different dithering functions [6].

For the quantizer output, we can write

$$Q\left(u_d(n,k)\right) = u_d(n,k) + q\left(u_d(n,k)\right). \tag{5.11}$$

After averaging

$$Q_{av}\left(u_d(n,k)\right) = u_{av}(k) + \frac{1}{N_E}\sum_{n=1}^{N_E} q\left(u_d(n,k)\right). \tag{5.12}$$

Assume that the dithering signal is chosen such that it is *self-subtractive*, meaning that its average over the entire ensemble is zero, which is the case for both of the dithering functions above. Then we have

$$u_{av}(k) = u(k) \tag{5.13}$$

$$Q_{av}\left(u_d(n,k)\right) = u(k) + q_{av}\left(u_d(n,k)\right) \tag{5.14}$$

$$q_{av}\left(u_d(n,k)\right) = \frac{1}{N_E}\sum_{n=1}^{N_E} q\left(u_d(n,k)\right), \tag{5.15}$$

where q_{av} is the quantization noise of the averaging quantizer $Q_{av}(\cdot)$.

Note that noise dither is not a better choice for dithering in this technique, because we desire to have dithering functions that eq. (5.1) have completely zero average for limited ensemble size, even for $N_E < 10$, and eq. (5.2) can spread the dithered values over the given range more evenly. These two criteria are not always met if noise dithering is used in a limited-size ensemble.

In the rest of this section we show the ensemble dithering results applied to a quantizer. The input to the quantizer has values in the range $[-0.5, 0.5)$. As mentioned earlier, limiting the input within this range does not lack the generality of the method, since the quantization mechanism repeats periodically by adding one unit to the input. The input range is divided into infinitesimal pieces and each resulting number is given to the quantizer to get its quantized value. The same procedure is repeated for the ensemble-dithered input. Then the mean square error (MSE) of all values is calculated for investigation. The ensemble size is five for all cases.

Figure 5.5 is a comparison of the quantization noise of an ideal quantizer with inputs of non-dithered $u(k)$ and two averaged $u_{av}(k)$'s with different dithering functions. A quantitative comparison MSE and MSE improvement relative to the non-dithered case can be seen in Table 5.1. The plots and the table show the significant reduction of quantization error by use of dithering in an ensemble.

Table 5.1: Improvement level with different dithering functions

Dither type	MSE	Improvement (dB)
No dither	0.289	0
Ramp	0.058	14.0
Sine	0.063	13.2

5.3.2 Optimal sine dithering

As can be seen in eq. (5.7) and eq. (5.8), the ramp dither function has no parameters to be optimized, whereas the sine dither function has A_m and Φ, which must be optimized for the best performance. Note that, in many applications, it is better to keep Am as small as possible to ensure that maximum dynamic range for the desired waveform is achieved and no clipping will occur. Fig. 5.6 shows the level of improvement in MSE for different ensemble sizes for several different values of A_m with zero Φ. Interestingly, all different ensembles have a common local maximum around $A_m = 0.4$. The reason for this value of local maximum will be the subject of future work. However, the value of the optimum Am varies for different ensemble sizes. For instance, for $N_E = 9$, assuming $A_m < 1$, the best value for A_m is 0.9. Nevertheless, there are larger optimum values that are beyond the range $[0, 1)$.

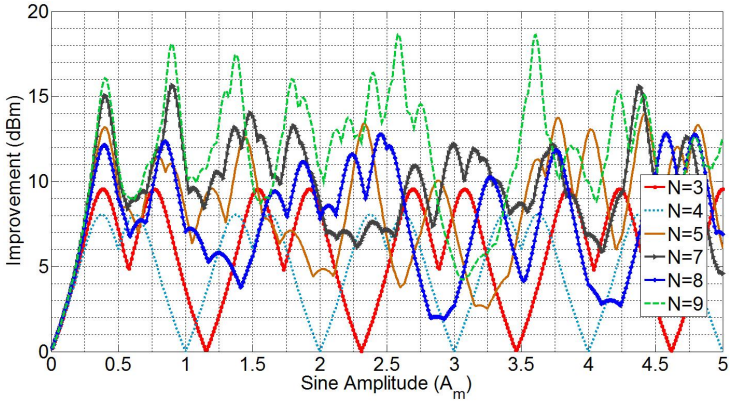


Figure 5.6: Simulated improvement in mean-square error vs. varying sine amplitude for various ensemble sizes N with zero phase Φ [6].

Figure 5.7 shows MSE improvement with both ramp and sine functions with various phase angles Φ versus ensemble size. From this figure, we see clearly that the ramp function always has better performance than that of the sine function. However, the sine function has comparable performance for odd ensemble sizes, especially below $N_E = 10$, with less than 1 dB difference from the ramp function improvement for optimized values of A_m and Φ. The "optimum" values presented here were obtained empirically. Development of analytic methods for determining the optimum values will be the subject of future work. Observations such as those in Fig. 5.7 led us to use sine dither for RF characterization, as illustrated in Section 5.4. Because sine dither is a continuous function and can be easily generated and measured in RF applications, we prefer this over ramp dither.

We see in Fig. 5.6 and Fig. 5.7 that odd ensemble sizes outperform even ones when the sine function has a zero initial phase, as in eq. (5.8). The reason can be explained in Fig. 5.8, in which a unit circle is illustrated. In a unit circle, the projection of a radial line with an angle onto the vertical sine axis determines the value of sine for that angle. The angles with thick lines are for $N_E = 3$, whereas the thin lines are for $N_E = 6$. This figure clearly shows that the ensemble size of six has duplicated dither values of the ensemble size of three; i.e., it has three redundant dither values that lead to no additional improvement compared to $N_E = 3$. Hence, in Fig. 5.6 the ensemble size of six has a curve that completely matches that of $N_E = 3$. The same holds for other even values of $N_E = 2(2k + 1)$, which have the same dither values of $2k + 1$. Therefore, the best choice for ensemble size is an odd value. However, if the

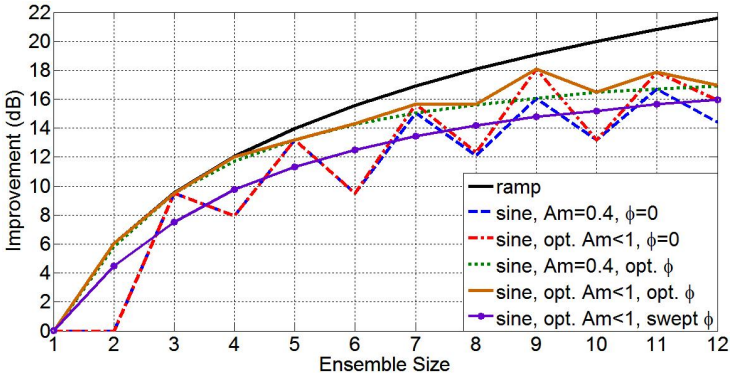


Figure 5.7: Simulated improvement in mean-square error vs. ensemble size for different dithering functions [6].

phase Φ of the dither sine function is optimized empirically, the even-number cases will also show an acceptable improvement. In this way, the projections of different segments have different values on the sine axis and the problem of redundancy is resolved. As a rule of thumb, the optimal value of Φ is around $\pi/2N_E$, although it is optimized more accurately later in this chapter. With this optimization, we obtain MSE improvements, as shown in Fig. 5.9. Note the enhancement of even-N cases, as compared to Fig. 5.6.

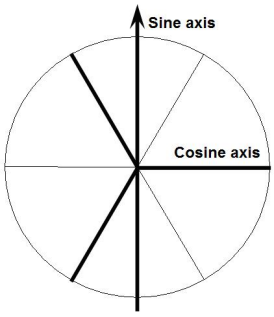


Figure 5.8: Different phase angles used in sine function dithering with zero initial phase. The thick line is for $N_E = 3$ and the thin line is for $N_E = 6$ [6].

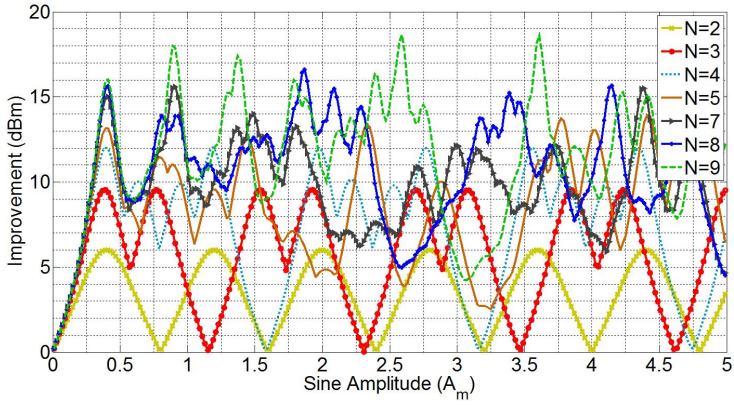


Figure 5.9: Simulated improvement in the mean-square error vs. varying sine amplitude for various ensemble sizes N with an optimized initial phase of the dithering sine function. Note the enhancement of even-N cases as compared to Fig. 5.6 [6].

5.4 Ensemble Implementation in RF Measurements

For the realization of the ensemble-averaged dithering approach discussed in Section 5.3 for RF test and characterization, we should find the best way to construct the required ensemble. Repeating a measurement several times is not the optimal approach, because the time between two consecutive data captures can be long enough for unwanted variations, though small, such as temperature and bias conditions, or for variations of the signal generator output over time. It is better to reduce the interval between two consecutive measurements to minimize the measurement uncertainty.

5.4.1 Periodic ensemble approach

We propose an efficient way for construction of the dithered ensemble, using only a single measurement. Assume that we need an ensemble size NE for dithering. In RF test and characterization we are often interested in periodic signals. Suppose that the input signal is represented as

$$x(k) = \mathrm{Re}\{\tilde{x}(k)exp(j2\pi f_c k)\}, \tag{5.16}$$

where $\tilde{x}(k)$ is the baseband signal with period T_{env} and f_c is the carrier frequency. Because the modulated-signal envelope period is not always an integer multiple of the RF carrier period $T_c = 1/f_c$, we define T, the modulated-signal repetition period, as the least common multiple of T_{env} and T_c. That is,

$$T = \text{LCM}(T_{env}, T_c), \tag{5.17}$$

where T is the period of the modulated RF signal $x(k)$.

To illustrate T, T_{env} and T_c more clearly, they are shown in Fig. 5.10 for a periodic modulated waveform. Now, if we construct a dithered waveform as

$$x_d(k) = x(k) + d(k) \quad , \quad 0 \leq t \leq N_E T, \tag{5.18}$$

with the designated time interval $N_E T$, and $d(k)$ as the added dither, we will

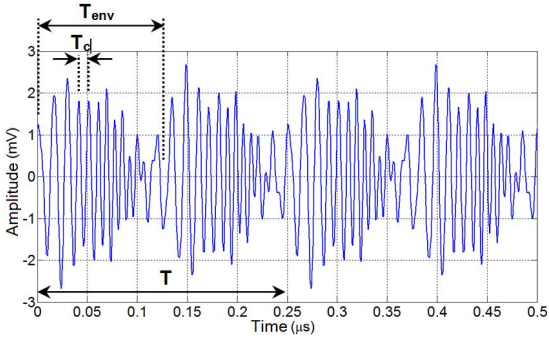


Figure 5.10: Illustration of different characteristic periods of a simulated 11-tone signal with bandwidth of 8 MHz and center frequency of 100 MHz [6].

show that this waveform can be used as a complete ensemble of size N_E for the original waveform $x(k)$ with the time interval $[0, T)$. We define our averaging method over $x_d(k)$ in the following way:

$$\begin{aligned} x_{av}(k) =& \frac{1}{N_E} \sum_{m=0}^{N_E-1} x_d(k+mT) \quad , \quad for \quad 0 \leq k < T \\ =& \frac{1}{N_E} \sum_{m=0}^{N_E-1} x(k+mT) + \frac{1}{N_E} \sum_{m=0}^{N_E-1} d(k+mT), \end{aligned} \tag{5.19}$$

with the designated time interval $[0, T)$. Note that in contrast to $d(n)$ in eq. (5.10), which was defined as a function of the sample number n, $d(k)$ in the

above equation is a function of time, which means that the whole ensemble is created in a single time sequence. However, for $d(k+mT)$, we assume the same self-subtractive properties as that of $d(n)$. Therefore, since $x(k)$ is periodic with period T, eq. (5.19) is simplified into

$$x_{av}(k) = x(k) \quad , \quad 0 \leq k < T. \tag{5.20}$$

Hence:

$$\begin{aligned} Q_{av}\left(x_d(k)\right) &= x(k) + \frac{1}{N_E}\sum_{m=0}^{N_E-1} q\left(x_d(k+mT)\right), \\ &= x(k) + q_{av}\left(x_d(k+mT)\right) \quad , \quad 0 \leq k < T. \end{aligned} \tag{5.21}$$

$$\begin{aligned} q_{av}\left(x_d(k+mT)\right) &= \frac{1}{N_E}\sum_{m=0}^{N_E-1} q\left(x(k) + d_t(k+mT)\right), \\ &= \frac{1}{N_E}\sum_{m=0}^{N_E-1} q\left(u(k) + d_t(k+mT)\right), \end{aligned} \tag{5.22}$$

with the dithering function $d_t(k')$ as either of the following self-subtracting functions:

$$\text{ramp}(k') = \frac{k'}{N_E T} \tag{5.23}$$

$$\text{sine}(k') = A_m \sin(2\pi k'/N_E T). \tag{5.24}$$

For the same reasons explained in the previous section, the use of a sine dither function is more desirable than the use of ramp function. Thus, a sine dither function was used in the following analysis. The time plot of an entire sine-dithered 11-tone multisine signal [49] with 4-bit resolution and an ensemble size of five, is shown in Fig. 5.11. In this figure, the center frequency is 4.5 GHz and the bandwidth is 80 MHz. Therefore, $T = 4T_{env}$, and $5 \times 4 = 20$ envelope cycles can be seen. The sine-shape lines on the top and the bottom of the plot show the amount of variation caused by the sine dither, which is a sine with amplitude of 0.4 times the least significant bit (LSB), $0.4 \times 62.5 = 25$ mV.

Figure 5.11 also visualizes how eq. (5.19) operates. Equation (5.19) states that in order to find the post-processed ensemble-dithered value of the waveform at a specific time instant k, we should take the average of the corresponding

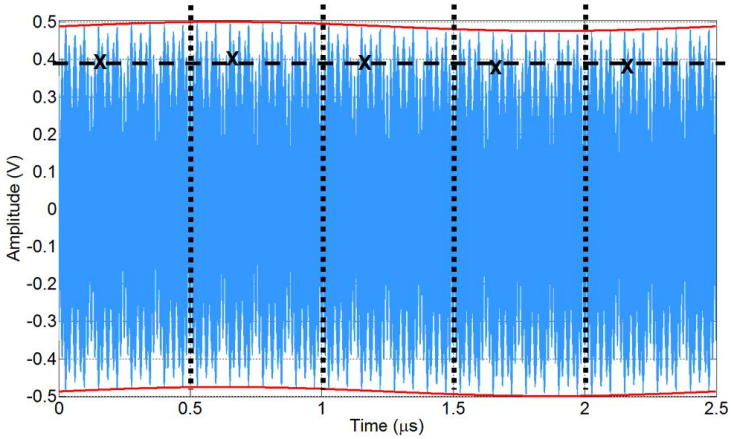


Figure 5.11: Time-domain plot of a simulated 11-tone signal with the proposed dithering technique and $N_E = 5$ [6].

values in different segments (with length T) of the repeated signal with the dither added. In Fig. 5.11, an example of such values are illustrated by cross symbols. They are corresponding points in different segments separated by vertical dotted lines. These points would not have been different in value if the waveform had not been dithered. However, due to dithering, their values differ a little, as can be seen by the small fluctuations around the horizontal dashed lines.

Note that this dithering method is designed to reduce quantization noise in DACs. Hence, the low-frequency dithering signal is not attenuated in the digital domain before passing through the quantizer, even though the system may have a band-pass output. However, the low-frequency component may become attenuated upon passing through the output frequency-selective circuits, which are after the quantizer. Such attenuation is not a problem because this low-frequency signal is not desired after quantization and it will be removed by averaging.

5.4.2 Verification of proposed approach

Comparing eq. (5.15) and eq. (5.22), if we consider each time sample individually, we may conclude that the averaged quantization noise versus u is the same as the one shown in Fig. 5.7 for the two dithering functions, ramp and sine.

However, for the averaged MSE over time, we must make two assumptions to be able to compare to the static case. First, the sine dithering function should have the same phase as that of the static case. However, because the argument of the function d in eq. (5.22) is $k + mT$, the sine dither has a nonzero phase of $(2\pi k/N_E T)$, and it varies with k. In order to compare the former static case with the current dynamic case, we modify our analysis of the static case as:

$$\text{sine}(n, p) = A_0 \sin(2\pi n/N_E + 2\pi p/P), \tag{5.25}$$

$$\text{where } n \leq N_E \quad , \quad p \leq P,$$

with a new additional argument p, where n and p are integer numbers, and P is the maximum value of p. P should be a large number so that we ensure that enough phase samples are considered. The MSE values for every p are calculated as in Section 5.3. Then, these values are averaged over the entire p-value range of $[1, P]$. The resulting MSE improvement, as compared to the nondithered case, is added to Fig. 5.7 as a solid line with circle markers. Again, in this case, the optimum value for the sine amplitude A_0 is 0.4, as can be seen in Fig. 5.12. Note that for all of the ensemble sizes less than 12 the optimum A_0 value is 0.4, unlike Fig. 5.6 and Fig. 5.9. The amount of MSE improvement is less than that of the optimized sine case in Fig. 5.7, because the phase has been swept and the MSE value is averaged. However, the improvement is only around 2 dB less than the optimized case with different N_E. With this average MSE of the phase-swept sine dither, we ensure that the resulting static analysis is consistent with the analysis of the dynamic case.

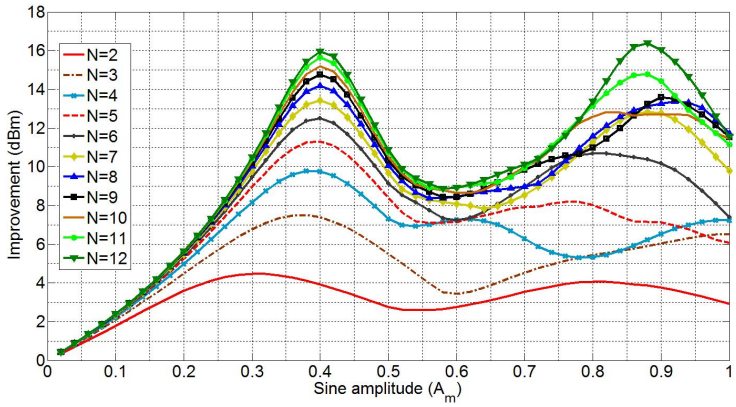


Figure 5.12: Simulated MSE improvement vs. varying sine amplitude and ensemble size with the proposed dithering method [6].

The second assumption required for a valid comparison of static and dynamic dithering is that the probability of occurrence of all u values is equal, which is a reasonable assumption if the waveform amplitude variation is much larger than the bit resolution Δ of the quantizer and the waveform is complex [106]. Hence, with the proposed method and in RF measurement situations where the waveform range is large as compared to Δ, and the waveform is a complex modulated signal, we have quite similar MSE values and improvement to that of the swept-phase case with the dithering function of eq. (5.25). This claim is verified in the following.

The test signal is a multisine waveform. Different ensemble sizes and different ADC resolutions were tested. The results are compared with the static case, as shown in Fig. 5.13. In this figure, N_b indicates the number of ADC bits. A very high level of agreement between the static and dynamic cases can be seen in this figure, which supports the analysis presented. The maximum difference between corresponding points of different curves is about 0.5 dB. Fig. 5.13 also shows that the number of bits in the data converter does not have a significant effect on the level of improvement. Although this is not shown in the figure, we have verified that the number of tones has no notable effect either.

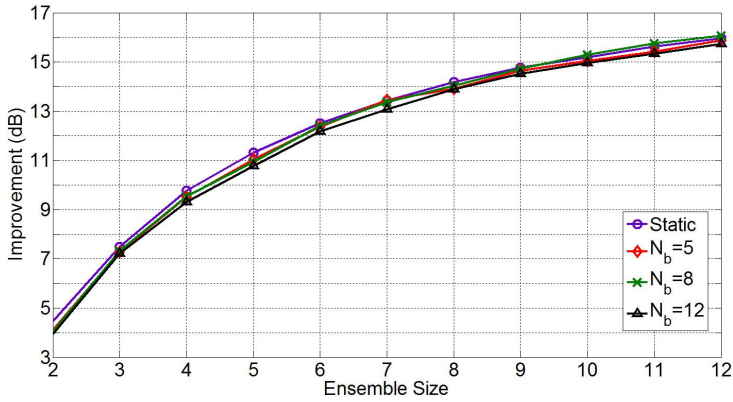


Figure 5.13: Simulated MSE improvement vs. ensemble size and number of bits with the proposed dithering method, as well as for the static case [6].

The simulated spectrum plots of a non-dithered and sine-dithered 11-tone waveform with 4.5 GHz center frequency, 160 MHz bandwidth and 20 GS/sec sampling rate are shown in Fig. 5.14 for an 8-bit resolution. The quantization noise reduction with the proposed method is clearly seen in this figure.

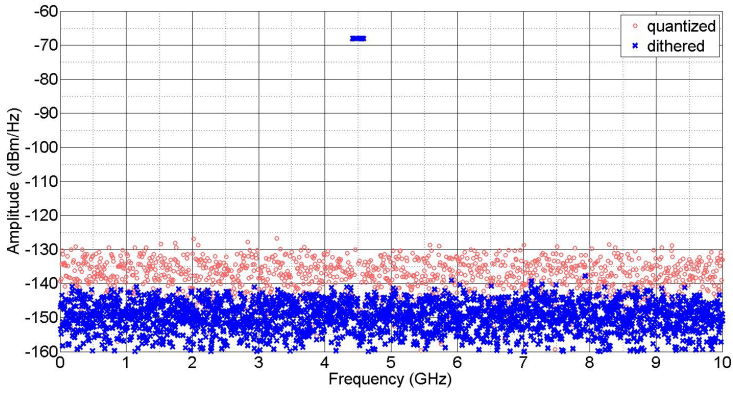


Figure 5.14: Simulated spectrum of the output of an ideal quantizer with an 11-tone input waveform for both non-dithered and dithered cases with the proposed technique [6].

5.4.3 Large dither for further non-ideality reduction

Real DACs have a transfer function that deviates from that of the ideal staircase shape. This deviation can be classified as a dynamic nonlinearity (DNL) or integral nonlinearity (INL) [107]. DNL effects can be reduced with dither, which is a well-known and widely used technique for improvement of data converters [108, 109].

In this section, we show how DNL is significantly alleviated by applying our proposed dithering technique, using a large dither signal. Assume that a 5-bit DAC deviates from its ideal staircase transfer function as shown in Fig. 5.15. Its transfer function has been distorted in 12 bins out of its 32 bins, as shown in the figure. This distortion causes some additional spectral impurity. Due to this DNL, the proposed dithering technique with the suggested optimum value does not work well because the DNL effects dominate the quantization noise.

However, by use of a larger dither signal, we can significantly reduce the DNL effects. This reduction is the result of delocalization and randomization of the sample location along the horizontal axis of Fig. 5.15 due to large dithering, which leads to different outputs with almost uncorrelated quantization error. Now, if the dithering amplitude is large enough to delocalize the sample to a location beyond the DNL points, averaging can give a value that is effectively close to the ideal case.

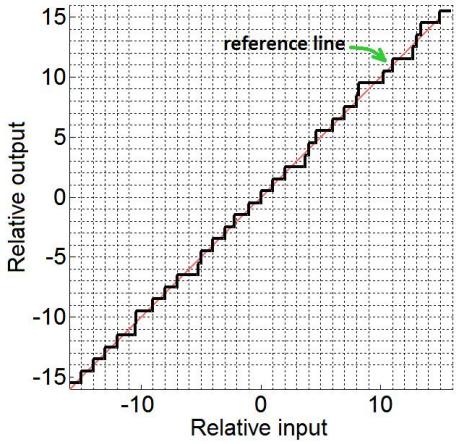
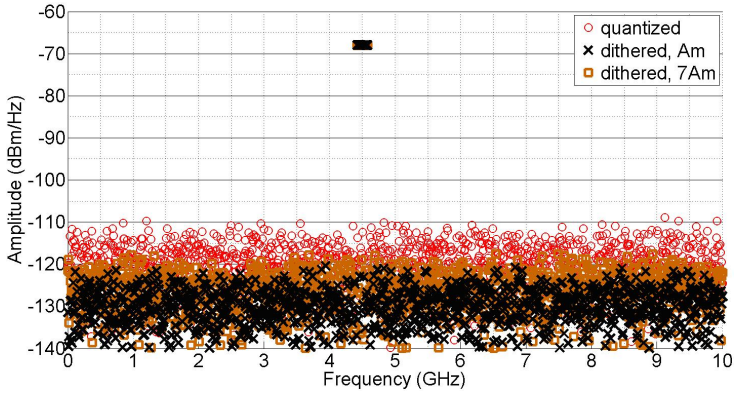


Figure 5.15: Simulated transfer function of a non-ideal 5-bit quantizer with dynamic nonlinearity introduced in 12 points [6].

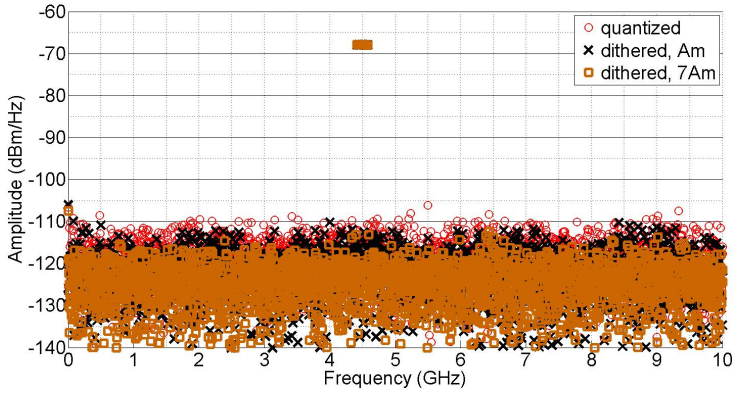
This mechanism is tested by applying different levels of dither. The results are shown in Fig. 12 and Table 5.2. For comparison, results in which the same dither level is applied to an ideal quantizer are also indicated. In Table 5.2, the criteria for comparison of these waveforms are ACPR and NRMSE. Due to the limited space, only the results of A_m and $7A_m$ cases are shown in the figures. Note that we define ACPR to include the increase in adjacent-channel power arising from both quantization noise and nonlinearity, as opposed to nonlinearity only, as is commonly done.

These results clearly show the effectiveness of large-amplitude dither in reducing the errors in a non-ideal DAC, assuming that the main cause of the non-ideality is DNL. Note that the optimum dither value in the ideal case is A_m, whereas $7A_m$ is the optimum in terms of NMSE in the non-ideal case. In the latter case, we are dealing with quantization error as well as an additional DNL error that needs larger dither for compensation. Also note that the average noise level in the latter case is higher than that of the former case for the same reason.

Note that INL effects cannot be reduced by dithering, because INL relates to the global variation of the transfer-function curve. Dithering can compensate only for local variations. Because we seek an efficient technique to reduce data converter non-idealities in real applications, based on the above discussion, we will choose large dithering for further investigation.



(a)



(b)

Figure 5.16: Simulated spectrum of the output of a 5-bit quantizer excited by an 11-tone input waveform for both non-dithered and dithered cases with the proposed technique and $N_E = 5$, for two dither amplitudes A_m and $7A_m$. a) The ideal case, and b) the non-ideal case with the dynamic nonlinearity given in Fig. 5.15. The dither amplitude Am is the optimum value as obtained in this chapter [6].

Table 5.2: Improvement level in post-processed simulated data of a 5-bit quantizer output with different dithering levels and $N_E = 5$.

Excitation	NMSE (dB)		$ACPR_{tot}$ (dB)	
	Ideal DAC	DNL DAC	Ideal DAC	DNL DAC
non-dithered	-29.1	-25.4	-47.2	-42.3
Dithered, 1x	**-40.0**	-28.7	**-58.1**	-44.7
Dithered, 3x	-34.3	-31.2	-52.5	-47.7
Dithered, 5x	-34.0	-31.2	-53.8	-47.8
Dithered, 7x	-36.3	**-31.3**	-53.3	-48.1
Dithered, 9x	-35.6	-31.1	-55.1	**-48.7**
Dithered, 11x	-33.6	-30.4	-58.1	-48.0

5.5 Measurement Results with Large Dither

In real applications, especially in the generation of RF waveforms through direct synthesis, there are several sources of nonlinearity that degrade the generated output. Some of them are briefly addressed in this section. One of the most serious sources of nonideality in RF generators is the ADC/DAC, although other non-ideal effects may manifest themselves as well. In this section, we show that we can efficiently use the proposed form of dithering to reduce nonidealities in the output from an arbitrary waveform generator RF source.

5.5.1 Measurement setup

The proposed method was applied to characterize the adjacent-channel distortion from an AWG. The AWG was programmed to generate a waveform with a modulation bandwidth of 160 MHz at a center frequency of 4.5 GHz. The setup used for this purpose is shown in Fig. 6.2. It consists of the following blocks, which are discussed in turn:

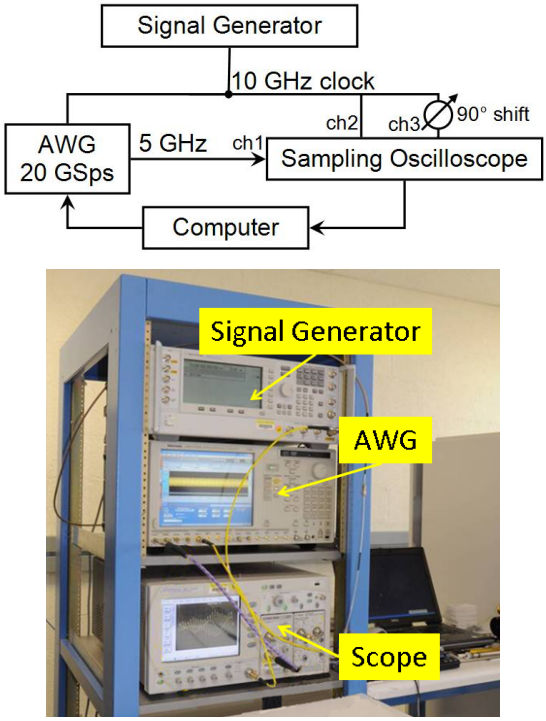


Figure 5.17: Experimental setup for test and characterization of an arbitrary waveform generator [6].

Arbitrary waveform generator

The AWG directly synthesizes RF waveforms with bandwidths up to several gigahertz by use of high sampling rates. For the present work, the sampling rate was set to 20 GS/s. As will be seen in Section 5.5.2, the AWG has some non-idealities that caused degradation of the generated output waveform.

Oscilloscope

We characterized the AWG with an oscilloscope whose samplers have a 50 GHz bandwidth. The oscilloscope provides equivalent-time sampling of periodic waveforms. The time step utilized here was 2.5 ps, corresponding to an effective sampling rate of 400 GS/s. We corrected the oscilloscope measurements in post

processing [110] for jitter, drift, and systematic errors in the oscilloscope's time base to accurately characterize the signal emanating from the modulated-signal source. Time-domain averaging was used to improve the dynamic range of the measurement to approximately 75 dB at 5 GHz. With this instrument we were able to detect different types of non-idealities, even at higher harmonics of the waveform.

Clock generator

For synchronization, a continuous-wave signal source provided a 10 GHz external clock signal for the AWG and was used for the time-base correction of the sampling oscilloscope measurements. This distribution of the 10 GHz reference is a key feature of the measurement setup [111], enabling precise synchronization of the various signals within the setup than could have been obtained through the use of a standard 10 MHz reference.

Excitation

To evaluate the proposed technique, a periodic multisine excitation was chosen as the test signal. It consisted of eleven tones with equal amplitudes, equal spacing of 16 MHz, and Schroeder relative phases [79]. The PAPR was 5.7 dB.

Measurement uncertainty

The measured results presented in this section focus on relative improvements in one measured signal with respect to another. Systematic errors common to all measurements will not significantly affect our results, so we focus on random errors represented by the standard deviation in our measured data. We report the highest value of standard deviation among the frequency components in an 11 tone, 160 MHz bandwidth, Schroeder-phase multisine centered at 4.5 GHz. The standard deviation was found by segmenting a nondithered waveform into $N = 13$ envelope cycles with the method described in eq. (5.19). We calculated the standard deviation of this set of measured data as

$$s_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}. \tag{5.26}$$

For this waveform, the maximum standard deviation in the magnitude was 0.03 dB, while the maximum standard deviation in the measured phase was 0.9 °.

5.5.2 AWG non-idealities

The AWG utilizes a DAC that directly synthesizes the RF signal from the digital data at a high sampling rate. In addition to quantization noise, other non-idealities cause the output signal to deviate from the ideal case. The most common of these non-idealities are addressed in the following.

Frequency dependence and memory

Like many RF circuits, the AWG has a nonideal frequency response. This phenomenon is more pronounced when the bandwidth of the waveform increases. For instance, Fig. 5.18 shows the measured spectrum of the AWG output, which was specified to be a multisine waveform with tones of equal amplitude. The measured amplitudes have been distorted due to the frequency dependent behavior of the AWG.

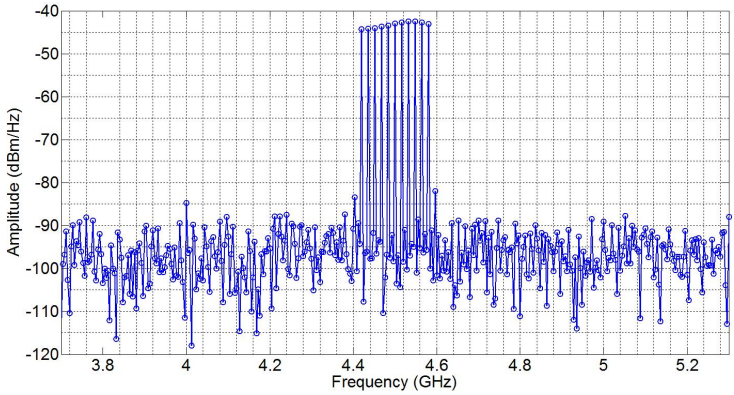


Figure 5.18: Measured output spectrum of an arbitrary waveform generator for an 11-tone multisine waveform [6].

Memory effects can also contribute to frequency dependent behavior. Memory effects cause the output of a system to be dependent on not only the present

input, but also the input values of the past. Various sources of memory effects may be classified as long-term and short-term memory effects [112].

The measured cumulative effects of different sources of memory for the AWG are plotted in Fig. 5.19. These are plots of the imaginary part vs. the real part (RI-plot) of the baseband input and output waveforms with modulation bandwidths of 50 MHz, 100 MHz and 1.25 GHz. If the system was memoryless, the RI-plot of the output would be a single curve, indicated as 'Target' in the figure. However, hysteresis effects can be easily seen in the plots of the output waveform, making it a double curve. As can be seen, a larger bandwidth results in larger hysteresis effects, indicating increased memory.

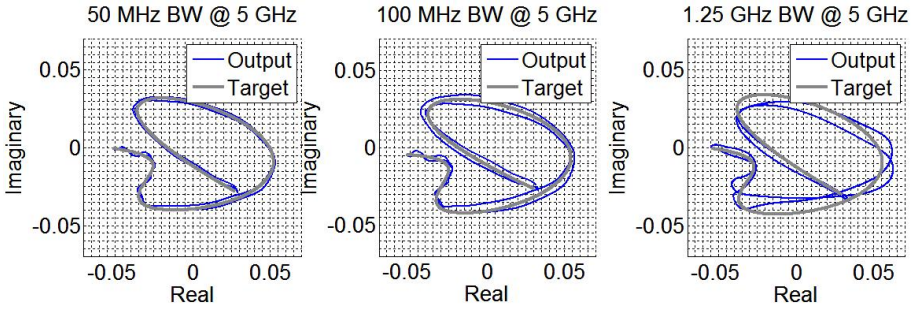


Figure 5.19: Measured output RI-plot of AWG for an 11-tone multisine waveform with bandwidth of (a) 50 MHz at 5 GHz, (b) 100 MHz at 5 GHz, and (c) 1.25 GHz at 5 GHz [6].

DAC Non-ideality

In an AWG, the main component that causes nonlinearity is the data converter. The DAC nonlinearity was already discussed in Section 5.4.3. This nonlinearity increases with operating frequency. The nonlinear transfer function of the data converter creates intermodulation products, as well as harmonics. The sampling nature of the data converter transfers the generated products to various frequency locations and spreads them throughout the spectrum. These spurs make the spurious-free dynamic range (SFDR) lower than expected. Another factor degrading the SFDR in DACs is a combination of glitches in the process of switching and timing mismatches [113].

Quantization Errors

As discussed above, this kind of non-ideality manifests itself in the time domain as a simple corrugation of the signal shape, whereas in the frequency domain we see this in the form of a large number of spurs in the out-of-band spectrum. If the quantization error is correlated with the signal, the number of spurs will be lower, but with higher amplitudes. Conversely, if the error is uncorrelated with the signal it manifests itself as a noise, resulting in a huge number of spurs with lower amplitudes.

The AWG is 8-bit in interleaving mode. Therefore, the signal-to-quantization-noise ratio (SQNR) should be around 48 dB [106]. Therefore, considering the signal bandwidth of 160 MHz and the sampling frequency of 20 GHz, the SFDR of the quantized spectrum should be 69 dB. Nevertheless, the measured output waveform has a lower SQNR, as shown in Fig. 5.18. This means that non-idealities of the AWG play important roles in SFDR degradation.

Different dithering levels based on the method proposed in this chapter were tested to study their improvement in SFDR. The measured results with $N_E = 4$ are shown in Fig. 16, and the corresponding ACPR values are illustrated in Table 5.3. Fig. 16 shows a reduction in the main band absolute power level with large dither, and this is due to the input normalization applied in the AWG.

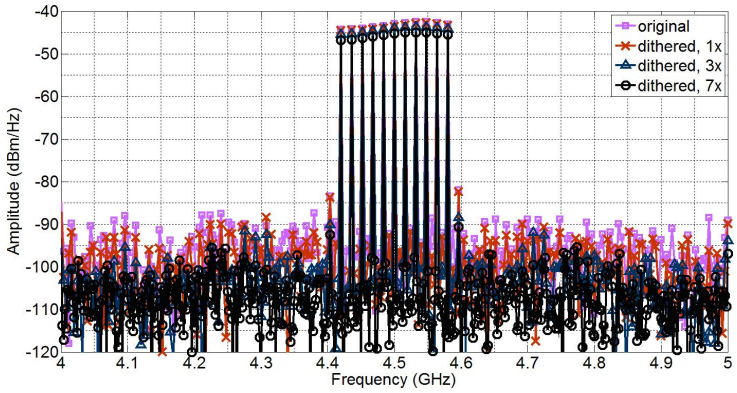


Figure 5.20: Measured output spectrum of AWG for an 11-tone waveform dithered with the proposed technique and $N_E = 4$, with different dithering levels [6].

Table 5.3: Improvement level in post-processed data of AWG output with different dithering levels and $N_E = 4$.

Excitation	$ACPR_L$ (dB)	$ACPR_R$ (dB)	$ACPR_{tot}$ (dB)
non-dithered	-47.9	-47.6	-47.8
Dithered, 1x	-47.9	-47.1	-47.4
Dithered, 2x	-49.7	-49.2	-49.4
Dithered, 3x	-51.0	-50.8	-50.9
Dithered, 5x	-52.7	-51.9	-52.3
Dithered, 7x	-53.9	-52.2	-53.0

The figure and the table show that the proposed dithering method improved ACPR by around 5 dB for the 160 MHz wide signal that was studied. As explained earlier and illustrated, for example, in Table 5.2, the real improvement is smaller than what is expected in the case of an ideal DAC due to non-idealities. We emphasize that the proposed method can be efficiently used in a very short time because it is based on a single measurement rather than many repeats.

Based on Sections 5.3 and 5.4, the improvement in SFDR by use of dither is mainly achieved through two mechanisms: (1) SQNR improvement and (2) effective DNL reduction. Note that dithering may also help in the reduction of hysteresis effects of the data converter building units, resulting in the reduction of memory effects [114].

5.6 Conclusion

In this chapter, the ensemble averaging dithering technique was introduced. The technique available in the literature uses averaging over different measurements each dithered with a different dithering noise signal. This technique requires the condition that the system characteristics do not change during different measurements. Moreover it requires a time much longer than a single measurement run. We developed a more efficient form of ensemble dithering for the case of periodic signals. This uses a simple deterministic sine wave in a single measurement to implement an efficient form of dithering by ensemble averaging in post-processing. The sine wave has a low frequency equal to the inverse of the measurement period. Therefore the measurement time will be

drastically reduced, since all different measurements in ensemble averaging will be integrated into a single measurement.

We applied this method to measurement of a modulated signal having 160 MHz modulation bandwidth at a carrier frequency of 4.5 GHz. We were able to reduce the adjacent channel distortion created by the signal generator by 5 dB.

Chapter 6

Cancellation of Receive-Band Noise in RF Transceivers

The contents of this chapter is partly based on the author's paper submitted to the IEEE Transactions on Microwave Theory and Techniques [7].

6.1 Introduction

With the advent of high speed communication and increasing demand for bandwidth, some new challenges are emerging. Recent telecommunication standards use higher bandwidth and, in some popular cases, small duplex spacing between the transmit (Tx) and receive (Rx) bands. Wide bandwidth and small spacing have some undesired consequences. An important instance of such an unwanted result is in RF transceivers, in which the transmitter and receiver operate at two different frequencies, and are electrically separated by an antenna duplexer with limited isolation. Spurious signals are typically produced by the transmitter in the Rx band. These spurious signals are caused by both *deterministic* and *stochastic* effects. The Rx-band unwanted signals, called Rx-band noise (RxBN) henceforth, may be non-negligible even after rejection by the duplexer. These signals leak into the receiver and can desensitize it, particularly in the case of large signal bandwidth and small duplex spacing. It should be noted that the RxBN discussed here is different from the second order intermodulation (IM2) product effect of the Tx leakage that occurs after

downconversion. The RxBN exists even before downconversion, right after the transmitter power amplifier (PA), due to the spectral regrowth and stochastic noise caused by the PA. This chapter explains RxBN effects, and presents some approaches to model and compensate them.

Section 6.2 decribes a typical scenario with the RxBN problem. In Section 6.3, we propose some techniques for characterization of RxBN. A novel method is also proposed to identify the nonlinearity orders that are the top contributors to RxBN. In Section 6.4, a Volterra-based modified memory polynomial is developed to model the deterministic portion of the RxBN, and an efficient method to prune the model coefficients while preserving the accuracy of RxBN is presented. Section 6.5 discusses some challenges in the RxBN modeling process. Some modifications to improve the RxBN model is proposed in Section 6.6. Since, as already explained, the RxBN problem cannot be completely prohibited in PA design, it should be cancelled in another way. Section 6.7 discusses two approaches to cancel RxBN, in which RxBN modeling can also be employed. Finally, the chapter highlights are summarized in Section 6.8.

6.2 Problem statement in a typical scenario

In current cellular deployment scenarios, the receiver and power amplifier are designed so that, along with duplexer isolation, the receiver sensitivity degradation is kept within acceptable limits – typically below 0.5 dB. However, future deployment of wider bandwidth and smaller frequency separation between transmit and receive bands, particularly in carrier aggregation mode, the emissions from the transmitter can significantly desensitize the receiver. These issues are further exacerbated by the need for lowering the supply voltage of the power amplifiers, in average power tracking [39] or envelop tracking [38] modes, to improved efficiency.

In order to estimate the influence of RxBN on system performance, consider a typical case in handsets for the Long Term Evolution (LTE) standard. The receiver input noise leaking through the duplexer at the receiver input should be less than -184 dBm/Hz (10 dB below the noise floor of -174 dBm/Hz), to keep receiver desense below 0.5 dB. Considering 50 dB suppression by the duplexer, the generated RxBN at the PA output should be less than -134 dBm/Hz. A typical handset PA has a maximum output power around 24 dBm at the antenna, which is equivalent to power spectral density of -46 dBm/Hz for a 10 MHz LTE signal. Therefore, the RxBN should be at least 88 dBc, which is challenging for certain cellular bands. For example, as visualized in Fig. 6.1, 10 MHz LTE

signal in band 13 with duplex spacing of 31 MHz, can easily fail the desired 88 dB suppression due to spectral regrowth caused by the non-linearity in the PA.

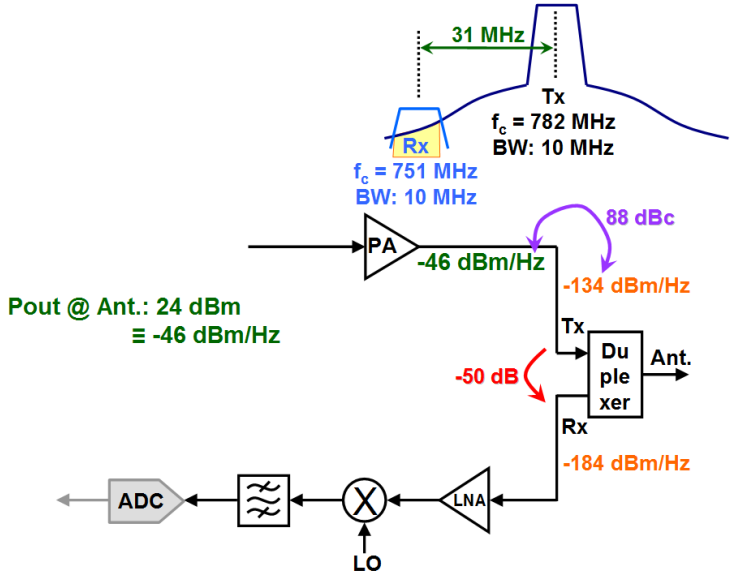


Figure 6.1: RxBN problem in a typical scenario of band 13 for LTE.

6.3 Experimental investigation of RxBN

The experimental setup used for RxBN investigation is shown in Fig 6.2. A waveform is generated in the Tx channel in the vector signal generator (VSG), and then filtered to remove the unwanted frequency components in the Rx band. The filtered signal feeds the PA, and the PA output which has frequency components in both the Tx and Rx channels passes through a surface acoustic wave (SAW) filter centered at f_{RX} and cascaded with a balun because of its differential output. The filter is actually a duplexer inserted in a reverse configuration such that the antenna port is connected to the PA, the Tx port is terminated with a 50 Ω load, and the Rx port is connected to the vector signal analyzer (VSA). In this way the PA is properly terminated, and the VSA effectively receives the RxBN signal because the Tx signal is highly attenuated at the duplexer Rx port. The RxBN signal detected by the VSA is transferred

to a computer (PC) as in-phase and quadrature (IQ) data for further signal processing. The VSA and VSG both work with 45 MHz sampling rate.

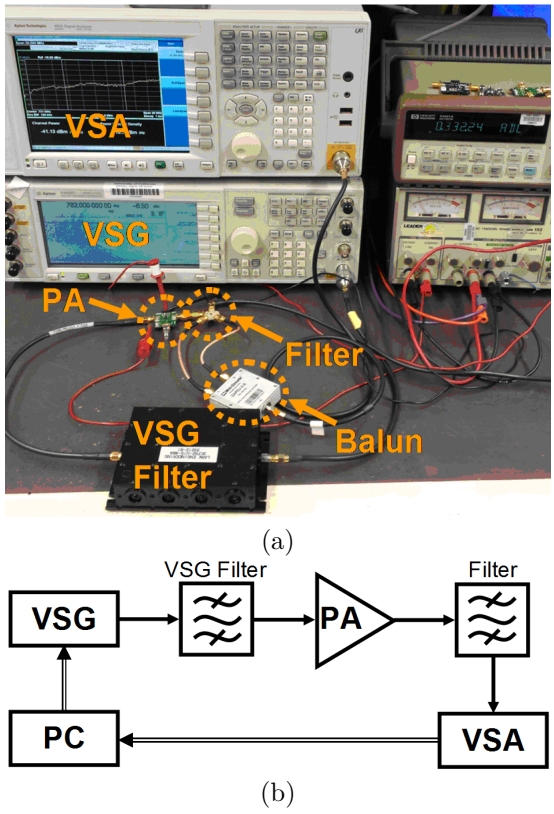


Figure 6.2: Picture and block diagram of the experimental setup [7].

Two commercial handset PAs (PA1 and PA2) from different manufacturers with the same maximum LTE output power of 27.5 dBm were tested. In our investigation, we use signals with 10 MHz bandwidth and 782 MHz center frequency, which is the uplink (Tx) band of band 13 for LTE. The corresponding downlink (Rx) band is centered at 751 MHz (-31 MHz offset from the Tx frequency). Two types of modulated signals were tested: an 11-tone multisine and an LTE excitation. Using multisine in investigations is beneficial, because it is much simpler than an LTE signal and helps us to understand some properties of the RXBN and the PA nonlinear behavior, as will be discussed next.

6.3.1 Measuring the deterministic components of RxBN

RxBN is a combination of *deterministic* and *stochastic* signals. The deterministic portion of the RxBN is generated by the nonlinearities of the PA which involves frequency translation from the Tx band to the Rx band. On the other hand, the stochastic part originates from random behavior of the PA. Two methods are presented in this subsection to distinguish deterministic RxBN from stochastic RxBN at system level. The first one which uses multisine excitation can be used in real-time assessment, while the other can be used for assessment in post-processing, but is more general.

Figure 6.3 shows the measured spectrum of PA1 in the RX band, including RxBN of the PA, at the average TX output power of 24 dBm. PA1 was excited by an 11-tone multisine in the Tx band with 10 MHz bandwidth. The figure shows two types of RxBN. One part of the RxBN is a set of discrete tones. Considering their frequency spacing and locations we deduce that they are deterministic noise component caused by nonlinearity. The other component is stochastic noise and manifests itself as continuous spectrum. Note that the observed stochastic RxBN is from the PA and is more than 10 dB above the noise floor of the instrument. The frequency scale in Fig. 6.3 is referenced to the TX center frequency. The observed slope in the continuous spectrum corresponds to a low roll-off of the duplexer, the matching network and the noise generated in the PA with increasing separation from the TX band.

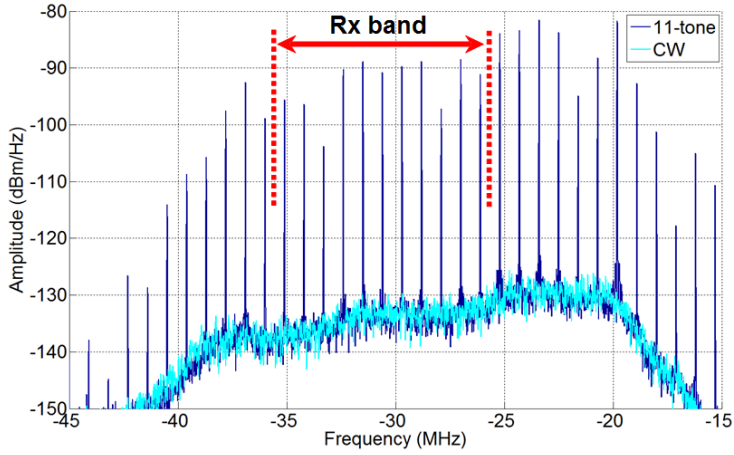


Figure 6.3: Measured RxBN spectrum of PA1 excited by an 11-tone multisine and a CW signal with the same frequency and power. The frequency axis shows the offset from the Tx frequency [7].

Table 6.1: Stochastic part of measured RxBN for multisine (MS) and CW excitation versus PA output power (P_{out})

P_{out} (dBm)	18	20	22	24
P_{nd} (dB), MS	-121.0	-115.9	-109.8	-106.8
P_{ns} (dB), MS	-136.9	-135.8	-134.9	-133.9
P_{ns} (dB), CW	-136.7	-135.7	-134.9	-134.0

By using multisine excitation and carefully choosing appropriate phase relationship between the tones, a peak to average power ratio (PAPR) of 8.5 dB, comparable to LTE signal was achieved. As shown in Fig. 6.3, RxBN has a major deterministic component evidenced by the tones that are well above the noise floor. We define the *determinism quotient* of the noise as:

$$D = \frac{P_{nd}}{P_{ns}} \tag{6.1}$$

where P_{nd} is the power of the deterministic component and P_{ns} is the power of the stochastic component of the RxBN. For the noise in Fig. 6.3, the determinism quotient is calculated to be $D = 27.1$ dB (P_{nd} is more than a hundred times larger than P_{ns}).

The observed stochastic noise originates from random processes in the PA, and its power depends on the average power of the PA. The RxBN generated by a CW excitation is illustrated in Fig. 6.3 against the discussed multisine with the same frequency and average. Table 6.1 compares the calculated Pns of measured data for 11-tone and CW excitations at different PA output powers. Fig. 6.3 and Table 6.1 show that the stochastic noise power is mainly determined by the average power, because it is the same for both multisine and CW excitations with the same output power.

Note that the roll-off at frequency offsets below -40 MHz and above -20 MHz is due to the limited capturing bandwidth and internal filtering in the VSA.

The multisine technique can distinguish between the deterministic and stochastic RxBN in real time. It uses the property of having discrete spectral contents with multisine excitation to discriminate between the deterministic and stochastic components of RxBN.

Measured RxBN with LTE excitation is shown in Fig. 6.4. Unlike CW or multisine, LTE excitation results in continuous RxBN spectrum, where the deterministic and stochastic components are indistinguishable.

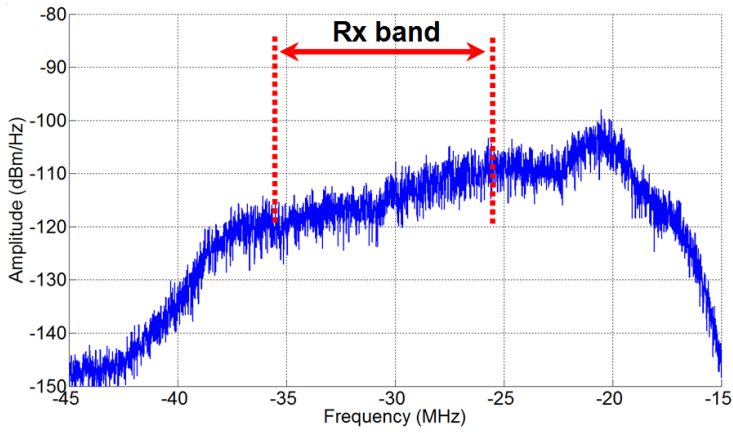


Figure 6.4: Measured RxBN spectrum of PA1 excited by an LTE excitation with the same frequency and power as those of Fig. 6.3. The frequency axis shows the offset from the Tx frequency [7].

Another technique that is capable of discriminating deterministic and stochastic RxBN even in continuous-spectrum cases, can be employed based on post processing the captured output signal. We call it the *repetition-subtraction* technique. In this approach, we capture more than one cycle of the signal, as illustrated in Fig. 6.5. Assume the output RxBN is $y(t)$, the period of the input signal $x(t)$ is T and we have captured an output signal of length $2T$. Also assume that Y_1 is the output sequence in $t \in [0, T)$ and Y_2 is the sequence in $t \in [T, 2T)$. Due to the periodicity of $x(t)$, if the RxBN is all-deterministic then $Y_1 = Y_2$ and $Y_1 - Y_2 = 0$. Hence, if Y_2 is subtracted from Y_1, the result will be only the stochastic residual RxBN signal (with standard deviation increased by 3 dB) because the stochastic part is not periodic. Therefore, by this post-processing technique we are able to quantitatively measure how much of the RxBN is deterministic and how much is stochastic. The advantage of this approach is that it can be applied irrespective of the waveform modulation, and can thus be used even for continuous-spectrum signals.

Figure 6.6 shows the post-processed stochastic noise against the total RxBN of different modulations, all measured at the average power around 24 dBm. Like Fig. 6.3 and Table 6.1, this figure also shows that the power of the stochastic part of the RxBN is mainly determined by the average power and is almost independent of the signal dynamics.

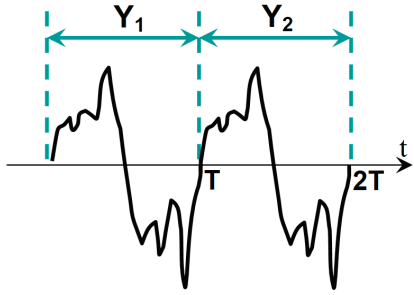


Figure 6.5: Repetition of signal, used for repetition-subtraction technique [7].

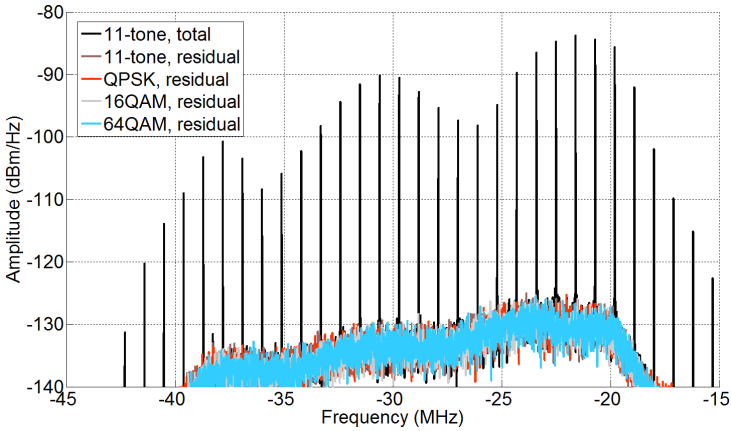


Figure 6.6: Total RxBN of measured data from PA1 and its stochastic RxBN for different modulations, calculated with repetition-subtraction technique. The frequency axis shows the offset from the Tx frequency [7].

6.3.2 Identification of nonlinearity orders

Before starting to set the model parameters for identification, it is helpful to get a priori knowledge of the strength of nonlinearity orders. In this section we propose a novel method for the assessment of nonlinearity order strength contributing to the RxBN signal.

The method is based on offset multisine excitation. Offset multisine has been recently proposed as a strong tool to characterize different intermodulation terms lying in and around the main band [5]. The technique is based on offsetting the frequency locations of the tones in a conventional multisine such that the

intermodulation (IM) products lie at frequencies different from each other in order to be individually distinguishable. However, the proposed technique cannot be used for receive-band characterization in its original form, and some simplifications have to be made to apply it to RxBN. Three-tone offset multisine was used, and the memory effects were ignored for the moment, as our primary goal was to assess the strength of the different nonlinearity orders qualitatively.

The frequencies of the three tones are $f_1 = f_c - f_m$, $f_2 = f_c$ and $f_3 = f_c + f_m - \delta$, in which f_c is the center frequency, f_m is the major frequency spacing between the tones and δ is the offset of the third tone from its original location as compared to the conventional three-tone. If a conventional multisine excitation is used, the deterministic RxBN will be a set of tones with f_m spacing. However, if the third tone is offset by δ there will be a number of tones scattered with δ spacing accumulated around the frequencies spaced by fm. Figures 6.7a and 6.7b show the RxBN of PA1 excited by a conventional and an offset three-tone multisine, respectively. In these figures f_c = 782 MHz, f_m = 5 MHz and δ = 20 kHz. The resulting deterministic RxBN consists of IM products caused by different nonlinearity orders. The IM products falling around the fundamental frequency, including those in the receive band, are caused by odd-order nonlinearities. In general, the frequencies of these IM products are:

$$f_{IM} = mf_1 + (1 - (m - n))\, f_2 - nf_3, \tag{6.2}$$

where m and n are integer numbers. The resulting offset from the center frequency is $f_\Delta = -(m + n)f_m$ for the conventional multisine case and $-(m + n)f_m + n\delta$ for the offset multisine case. Therefore, the frequency difference between the two cases is $n\delta$. Hence, if we measure the deviation of each tone from $-(m + n)f_m$, which is equal to $n\delta$, we can find the corresponding n-value. Having n and $f_\Delta = -(m+n)f_m$, we can calculate the corresponding m. Finally, based on eq. (6.2) the corresponding nonlinearity order becomes:

$$k = |m| + |n| + |1 - (m - n)|. \tag{6.3}$$

Applying this technique to the spectrum of Fig. 6.7b, the nonlinearity orders of the tones around -35 MHz offset from the transmit frequency are derived and illustrated in Fig. 6.7c. Note that the tones assigned to lower nonlinearity orders also have contributions from higher orders with various phases, but not vice versa. Nevertheless, ignoring memory effects we can assume that the contribution of, for example, the 17^{th} order to the tones assigned to the 15^{th} order is not larger than -10 dB, because the differences between the 15^{th} and the 17^{th} tones are about 10 dB. Therefore, the plot provides us qualitative assessment and comparison of the strength of nonlinearity orders, and helps to decide up to what order to consider for RxBN modeling.

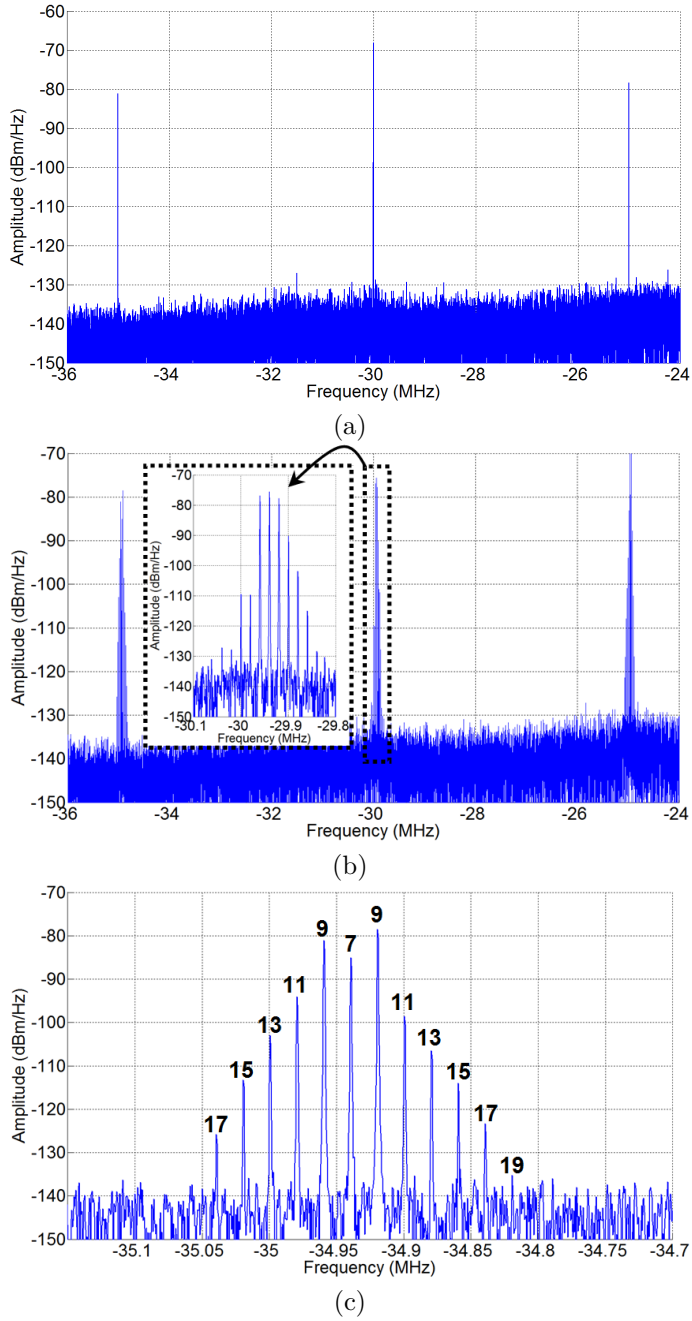


Figure 6.7: Measured PA1 output spectra excited by three-tone multisine with 10 MHz bandwidth: a) conventional multisine, b) offset multisine and c) zoomed view of spectrum for offset multisine with reduced stochastic noise by averaging. The nonlinearity order corresponding to each tone is indicated above it. The frequency axis shows the offset from the Tx frequency [7].

6.4 Modeling of Deterministic RxBN

Modeling RxBN is beneficial in some applications. It can be used for system-level analysis and evaluation where a fast but sufficiently accurate estimation of RxBN is required. By interpreting different terms of the RxBN model, one may achieve better understanding of the RxBN generation mechanism. Additionally, as will be seen in Section 6.7, it can be used for RxBN cancellation schemes. This section proposes a way to model RxBN.

RxBN modeling is different from the in-band modeling for error vector magnitude (EVM), since the frequencies of the input signal and the target signal are different. Also, unlike the modeling of adjacent-channel leakage power (ACLR), the RxBN is effectively a combined result of spectral regrowth caused by various high order nonlinearities in the power amplifier. Moreover, the RxBN is practically too far from the Tx-band to be estimated by a conventional behavioral model.

Recently, there have been a few attempts to model the RxBN of power amplifiers [42, 45]. In [42], the investigation was limited to only simulation, where the PA was assumed as a Wiener model consisting of a linear filter in series with a static nonlinearity. The RxBN was then estimated by using a parallel Hammerstein model. The paper did not discuss any real cases with experimental results and did not cover important considerations that arise in practical applications addressed in this chapter. In [45], experimental measurements from a power amplifier were considered. However, the steps required for successful modeling of the RxBN were not presented. Furthermore, there was no discussion of the dependence of the modeling accuracy on the power level of the RxBN, nor methods to distinguish deterministic from stochastic part of RxBN.

In traditional Tx-band modeling that uses Volterra-based behavioral models, one tries to model the PA output around fundamental frequency based on the input [115, 8, 116, 63, 2, 61]. More precisely, the frequency components of the PA output in the Tx band and its vicinity are modeled by using a Volterra-based model. These models usually yield acceptable estimation accuracy within a bandwidth up to three or five times the input bandwidth. The frequency components beyond this bandwidth are usually not modeled well. This shortcoming makes RxBN estimation challenging, due to the separation of the Rx band from the Tx band.

Signal frequency components from the main band translate to receive band due to nonlinearity of the power amplifier. The output of nonlinearity with the order N excited by a signal of bandwidth BW and with a center frequency f_c, is a signal with a frequency span from $f_c - NBW/2$ to $f_c + NBW/2$. If the

nonlinearity order is large enough, the output signal spectrum can spread into the receive band frequency, causing RxBN generation.

The modeling approach used in most of Volterra-based behavioral models for RF applications, including those in this chapter, is in baseband domain [57]. To model the abovementioned frequency translation mechanism, a modification to the conventional behavioral models is needed. For example, one of the most widely used models is the memory polynomial (MP) model [61]:

$$y(n) = \sum_{m=0}^{M-1} \sum_{k=1}^{K} b_{mk} |x(n-m)|^{2k-1} x(n-m), \tag{6.4}$$

where M and K are the memory depth and nonlinearity order, and $x(n)$ and $y(n)$ are the input and output of the circuit, respectively. The MP model is very effective for main-band modeling. However, it is inefficient for RxBN modeling, since it has some excessive low-order terms that do not contribute to RxBN. Therefore, for RxBN modeling, the MP model is modified as:

$$\begin{aligned} y(n) =& F\{ \sum_{m=0}^{M-1} \sum_{k=1}^{K} b_{mk} |x(n-m)|^{2k-1} x(n-m)\} \\ =& \sum_{m=0}^{M-1} \sum_{k=1}^{K} b_{mk} \cdot F\{|x(n-m)|^{2k-1} x(n-m)\}, \end{aligned} \tag{6.5}$$

where $F\{\cdot\}$ is a linear time-invariant (LTI) filtering operator that keeps only the frequency components which lie inside the receive band. The parameter k_0 is the minimum nonlinearity order contributing to RxBN. Based on what is explained in the first paragraph of this section, k_0 can be obtained by:

$$k_0 = \text{ceil}\left(|f_{TX} - f_{RX}|/BW + 0.5\right), \tag{6.6}$$

where ceil(x) is the *ceiling operator* returning the nearest integer number larger than or equal to x, and f_{TX} and f_{RX} are the transmit and receive frequencies, respectively.

The model is graphically illustrated in Fig. 6.8a. In this figure, z^{-1} is the delay operator, $P_n(\cdot)$ is a polynomial of an even order $2K$ with the input encountering an n-tap delay, $F\{\cdot\}$ is the LTI band-pass filtering operator with the center frequency at f_{RX} and w_n is the weight for $P_n(\cdot)$.

Each nonlinearity term in eq. (6.5) gives rise to spectral reshape and regrowth of the signal that can lie within the Rx band. After filtering by the $F\{\cdot\}$ operator,

only the Rx-band part of this signal remains and the rest will be removed. This mechanism is shown in Fig. 6.8b, where the signal spectra at different points of the nonlinearity path in the model are represented.

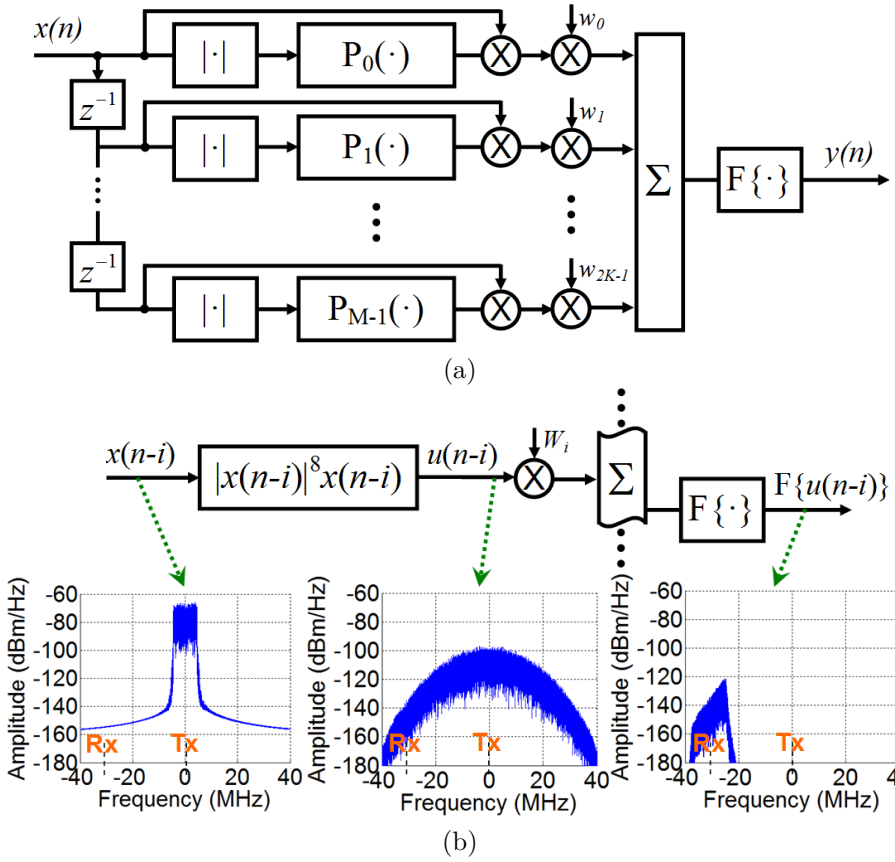


Figure 6.8: (a) Block diagram of the MP model used for RxBN modeling and (b) signal spectra at different points of the model path for the ninth-order nonlinearity [7].

6.5 Practical Challenges in RxBN Modeling

6.5.1 Signal Processing

Upsampling

Using eq. (6.5), it is clear that the frequency components resulting from the kth-order nonlinearity ($k_0 \leq k \leq K$) must not encounter aliasing in order to accurately model RxBN. Therefore, the sampling rate should be high enough to avoid aliasing. For signal bandwidth of 10 MHz and receive band at 31 MHz offset, the upsampling rate should be three to avoid aliasing up to the thirteenth-order nonlinearity.

As will be discussed in Section VI.A, even orders that will be used in the model generate a spectrum wider than the Nyquist zone. These even orders as well as odd orders higher than 13 encounter aliasing. However, the upsampling rate is chosen high enough so that the aliased components are much smaller than the original components in the Nyquist zone.

Filtering

Since the modeling goal is to predict deterministic RxBN waveforms as close as possible to the measured one, we need to filter the measured and modeled signals and keep only the receive band part for comparison. However, ideal filtering is not possible and some amplitude and phase distortion is inevitable. Since one of the major applications of RxBN modeling is real-time cancellation in RF transceivers, we use infinite impulse response (IIR) filters to have a compact implementable architecture. On the other hand, IIR filters with good pass band amplitude response have large distortion in their group delay response around the cutoff frequencies in the form of large peaks. In order to place the peaks' locations out of the passband, we have to increase the filter bandwidth. However, it should be noted that large bandwidth degrades modeling accuracy, because for this case the model tries to predict the signal in a bandwidth larger than the desired one. We use 12 MHz bandwidth which keeps the peaks outside the band of interest while reducing the accuracy by around 0.1 dB at 24 dBm of PA output.

Time alignment

Before starting model identification, the input and captured RxBN signals should be time-aligned to one another. Time alignment in RxBN modeling is not as straightforward as that of main-band modeling. In time alignment for main-band modeling, one aligns two signals: an input signal and an output signal very similar to the input with some small noise and/or distortion added to it. However, in RxBN modeling the input and the signal to-be-modeled, i.e., RxBN, are much different in terms of dynamics and frequency location. In other words, there is not any component in RxBN which is linearly related to the input.

The proposed method in this chapter for RxBN time alignment is to first pass the input signal through a nonlinearity and then filter the resulting signal with a passband filter around the Rx band. Since the deterministic RxBN is composed of nonlinear IM products the output of this nonlinear block can have some correlation with the RxBN. Therefore by taking the cross-correlation between the RxBN and the nonlinear block output, which is synchronous with the input, we can find the time location of best alignment. This alignment method is summarized in Fig. 6.9 . We have found that considering the modeling error as the figure of merit for time alignment, the best nonlinear block among those of the form $|x|^k$ is the one with $k = 1$.

For more precise alignment, subsample alignment [117] is also deployed in the process. In this case, interpolation in cross-correlation curve is used for subsample alignment. We investigated the effect of subsample misalignment, and found that the modeling accuracy is almost insensitive to subsample alignment. For example, the accuracy change is less than 0.1 dB for a misalignment of ± 0.5 sample, since the memory terms of the model can compensate for it.

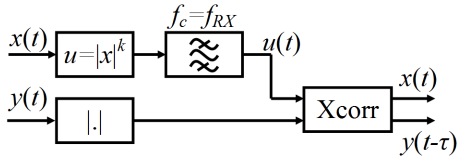


Figure 6.9: Block diagram of deployed time alignment [7].

6.5.2 Limiting factors

The model can estimate the deterministic part of RxBN. Therefore, if the stochastic part is large, the modeling accuracy will be limited. The available commercial PA designs have been optimized to keep the stochastic RxBN near or below the specifications (e.g., -134 dBm/Hz). Hence, the deterministic RxBN caused by spectral regrowth is usually the dominant noise.

As duplex spacing increases, the deterministic contributions are reduced, which makes the modeling more challenging and less accurate. However, in such cases lower power leaks into the receiver. Therefore, in large duplex spacing, the degraded model can still be acceptable enough to model the deterministic part.

6.5.3 Modeling sensitivity

The model of eq. (6.5) can be written in the matrix form as:

$$\mathbf{Y} = \mathbf{Xb}, \tag{6.7}$$

where $Y = [y(M)\ y(M+1)\ \cdots\ y(M+n)]^T$ is the output matrix, and $b = [b_{0,k_0}\ \cdots\ b_{0,K}\ b_{1,k_0}\ \cdots\ b_{1,K}\ \cdots\ b_{M-1,k_0}\ \cdots\ b_{M-1,K}]$ is the vector of coefficients to be estimated. The matrix $X = F[X_M\ X_{M+1}\ \cdots\ X_{M+n}]^T$ is the input matrix in which $X_p = [|x(p)|^{k_0-1}x(p)\ \cdots\ |x(p)|^{K-1}x(p)\ |x(p-1)|^{k_0-1}x(p-1)\ \cdots\ |x(p-1)|^{K-1}x(p-1)\ \cdots\ |x(p-M+1)|^{k_0-1}x(p-M+1)\ \cdots\ |x(p-M+1)|^{K-1}x(p-M+1)]^T$ and $F\{\cdot\}$ is the filtering operator defined in eq. (6.5). In estimating the model coefficients, eq. (6.7) is an overdetermined system of equations.

In real applications, $\mathbf{Y}$ is a noisy data stream and this noise can cause modeling performance degradation. Here, by noise we mean any disturbing signal other than the deterministic RxBN. This disturbance can be the stochastic RxBN, received Rx-band signal or additive noise after the PA, for example in the sampling circuit. In RxBN modeling, disturbance noise is more pronounced compared to the main-band modeling, because the signal to be modeled in RxBN modeling is much smaller than the main-band signal. Therefore, the disturbance noise, referred to as disturbance for brevity, is more comparable and can cause larger degradation. The sensitivity of eq. (6.7) to disturbance depends on the condition number κ of the input matrix $\mathbf{X}$. The condition number for rectangular matrices, including our input matrix $\mathbf{X}$, can be defined with respect to the 2-norm [58]:

$$\kappa(\mathbf{X}) = \frac{\sigma_{max}}{\sigma_{min}}, \tag{6.8}$$

where σ_{max} and σ_{min} are the maximum and minimum singular values of $\mathbf{X}$, respectively [58, 118]. This condition number reduces to the usual condition number when $\mathbf{X}$ is square. The larger the condition number, the more sensitive eq. (6.7) is to disturbance [58, 118].

In RxBN modeling, since we have to use a wide range of nonlinearity orders in eq. (6.7), the resulting $\mathbf{X}$ matrix has a large condition number which leads to higher sensitivity to disturbance.

In summary, sensitivity to disturbance is more pronounced in RxBN modeling as compared to main-band modeling due to two major reasons: 1) disturbance is more comparable to the signal to be modeled and 2) condition number is larger due to the wide range of nonlinearity orders required for modeling.

Considering model sensitivity to noise is important if we intend to use RxBN modeling in practice to mitigate the RxBN in a transceiver. For example, in an RxBN cancellation approach using RxBN modeling explained in Section 6.7.1 and Fig. 6.20, the model is trained by using a sequence of samples taken from the sampling circuit (auxiliary receiver). There is some inevitable noise added after the transmitter by the auxiliary receiver. If the RxBN level is comparable with this added noise, the sensitivity of the model will significantly influence the degree of cancellation. Furthermore, if we have a model robust to noise, we can relax the noise figure requirements of the auxiliary receiver to reach a more linear design (since there is typically a compromise between low noise and high linearity in the design of receivers).

In the next section we try to reduce sensitivity based on the condition number.

6.6 Modified MP Model

This section presents two steps to improve the RxBN model performance.

6.6.1 Including even-order terms

It was shown that including even-order terms can improve the performance of main-band modeling in the baseband domain [65, 66]. By even-order terms we mean the terms in the form of $x|x|^{2l-1}$ with l being a positive integer so that

the sum of exponents becomes an even number. Here we show that the same holds for RxBN modeling.

First we apply the proposed MP modeling of eq. (6.5) to a measured RxBN of PA1 at the average power of 24 dBm with a 10 MHz LTE excitation. The driven model, which includes only odd terms, is optimized based on the normalized root mean square error (NRMSE) defined by:

$$NRMSE = \sqrt{\frac{\frac{1}{N}\sum_{n=1}^{N}|\hat{y}(n)-y(n)|^2}{\frac{1}{N}\sum_{n=1}^{N}|y(n)|^2}}, \tag{6.9}$$

where $y(\cdot)$ is the measured value, $\hat{y}(\cdot)$ is the estimated value and N is the number of samples. The optimized parameters of eq. (6.5) are $M = 5$, $k_0 = 7$ and $K = 19$. The parameter values indicate the wide range of nonlinearity orders required for the optimal model, and indicate that different nonlinearity orders caused by different parts of the transmitter circuit contribute to the produced RxBN. The corresponding NRMSE value is -12.0 dB. The model coefficients here and in the rest of the chapter are identified by a least square (LS) method. Increasing the K value does not reduce the error. As shown in Section II.C, higher orders do not have significant effects on the RxBN of the PA under test.

One way to deal with this limitation is to include the even-order terms in the model. The resulting model is:

$$\begin{aligned} y(n) =& F\{\sum_{m=0}^{M-1}\sum_{k\in S_k} b_{mk}|x(n-m)|^{2k-1}x(n-m)\} \\ =& \sum_{m=0}^{M-1}\sum_{k\in S_k} b_{mk}\cdot F\{|x(n-m)|^{2k-1}x(n-m)\}, \end{aligned} \tag{6.10}$$

in which S_k is the optimal set of nonlinear orders that should be found. This set has both the odd and even orders. We apply the modified model in our estimation problem and find the optimized parameters. The chosen parameters are $M = 5$ and $S_k = \{2, 4, 6, 7, 9, 11, 13, 15\}$. The corresponding NRMSE value is -16.5 dB. Figure 6.10 shows the resulting spectra. In this modeling effort, as well as in the other RxBN modeling results in the rest of this section, the size of training and validation sequences were around 10000 and 420000 samples, respectively.

The performance of the obtained model is evaluated with changing PA1 output power, as illustrated in Table 6.2. As expected, the NRMSE magnitude is not

larger than D, because the model can estimate only the deterministic part of the RxBN signal. In the ideal case where the entire deterministic part is estimated, the residual error of the model is equal to the stochastic part. Hence, with an ideal model $|NRMSE| = D$. The last row of the table shows the residual error of the model. This error is close to -134 dBm/Hz, which is the higher bound of PA RxBN derived in Section I, so that the PA does not affect the receiver input noise floor. Therefore, the model performs well in this sense, and can be used in RxBN cancellation schemes.

Table 6.2 also shows the model performance with the same M and S_k values versus the power and RxBN properties of PA2 output. The model architecture is the same as that of PA1, but the coefficient values are recalculated for PA2. It is seen that the model performance for PA2 is not as good as that for PA1. However, the residual error is even better than that of PA1, because the total RxBN in PA2 was smaller than the total RxBN in PA1. It means that our modeling approach still performs well in terms of residual error, and can be used in RxBN cancellation schemes for PA2 as well as, or even better than, modeling for PA1.

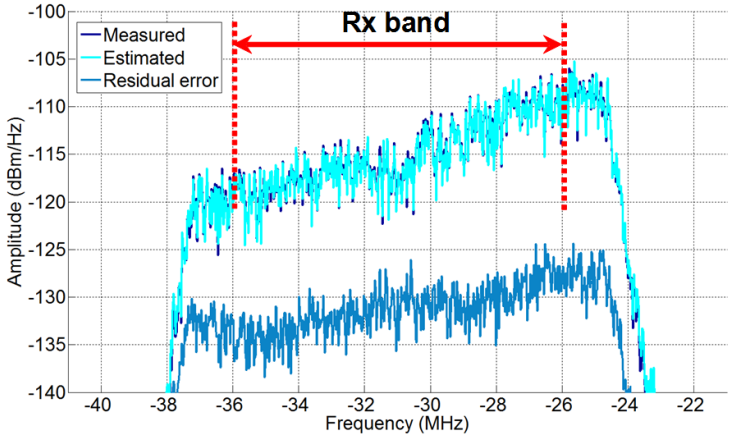


Figure 6.10: Spectra of the measured, estimated and residual RxBN for PA1 at the average output power of 24 dBm. The RxBN model is in the form of eq. (6.10) with $M = 5$ and $S_k = \{2, 4, 6, 7, 9, 11, 13, 15\}$.

We can see that by utilizing even-order terms we are able to get better performance. The use of even orders in the baseband domain is justified in [19], where they are used to model the dynamic bias and dynamic electro-thermal effects.

Table 6.2: Measured total RxBN, determinism quotient, NRMSE values of MP and power of residual error after using MP model versus output power (P_{out}) of PA1 and PA2

	P_{out} (dBm)	18	20	22	24
PA1	$P_{n,meas}$ (dBm/Hz)	-126.9	-122.4	-118.4	-113.5
	D (dB)	9.6	13.2	16.6	20.3
	NRMSE (dB)	-9.1	-12.1	-14.4	-16.5
	$P_{residual}$ (dBm/Hz)	-136.0	-134.5	-132.8	-130
PA2	$P_{n,meas}$ (dBm/Hz)	-129.6	-126.4	-123.4	-119.5
	D (dB)	13.1	14.2	17.3	20.3
	NRMSE (dB)	-10.8	-11.5	-12.2	-12.6
	$P_{residual}$ (dBm/Hz)	-140.4	-137.9	-135.6	-132.1

The spectral effects of baseband even-order terms (i.e., terms of the form $x|x|^{2k-1}$) are different from those of baseband odd-order terms (i.e., terms of the form $x|x|^{2k}$). Fig. 6.11 shows the spectra of the second- and third- order nonlinearities excited by a two-tone signal. This figure shows that the third order causes spectral regrowth in the region of $3BW$, while the second order causes spectral regrowth within an infinitely wide region. The same fact holds for higher orders.

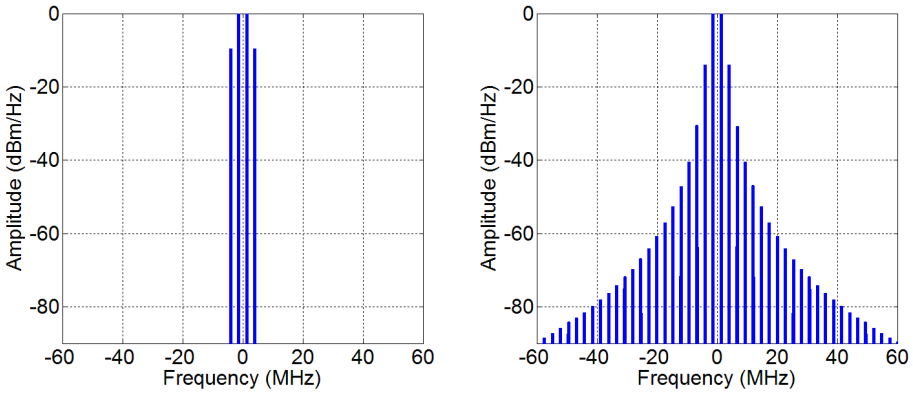


Figure 6.11: Spectra of a) third- and b) second- order nonlinearity excited by a two-tone signal [7].

The MP model performance at 24 dBm power level for PA1 was compared with

three other models: the dynamic deviation reduction (DDR) [63] model with some pruning, a simplified version of it given in [2] with some pruning, and the envelope memory polynomial model [62]. The three models were modified like the MP model in eq. (6.5), Section IV, to be able to model the RxBN. The parameters selection and the pruning in the three models were tried to be optimal. The corresponding NRMSE values for these models were -14.0 dB, -14.5 dB and -14.9 dB, respectively. Thus, the proposed MP model has higher performance than all of these three models, although it does not have cross terms.

6.6.2 Conditioning of input matrix in modeling

As mentioned in Section 6.5.3, the matrix equation of RxBN modeling problem is given by eq. (6.7), where $\mathbf{X}$ is the input matrix consisting of different input samples at different time instances and their different orders. This special type of matrix can have a large condition number explained by eq. (6.8), because its elements can be highly correlated together. A large condition number can result in high sensitivity to noise, as mentioned in Section 6.5.3. The matrix conditioning of $\mathbf{X}$ in eq. (6.7) is discussed in this subsection.

First, we examine the singular values of the input matrix which reflect the condition number. The singular values can be obtained by the singular-value decomposition (SVD) method [58, 118]. These values show how many dimensions of the matrix are important in calculating the matrix inverse. The singular vectors corresponding to small singular values show the directions of the vector space, or dimensions of the transformed vector space, which have insignificant contributions in determining the input matrix elements. These directions are the ones prone to perturbation, or in other words, sensitive to noise [58]. Figure 6.12 shows the singular values of the input matrix. As can be seen, they have a very wide range, and the ratio of the maximum to the minimum singular value is larger than 10^7.

In the next step, we use truncated SVD [58] for matrix inversion in model identification to show that the direction of the singular vectors corresponding to the small singular values have insignificant contribution to the input matrix. Truncated SVD can be used for sloving ill-conditioned LS problems, i.e., LS problems with large condition numbers. In this method, the small singular values and their corresponding singular vectors are first eliminated, which means that the effective dimensions of the problem are reduced, and then the matrix inversion is made in the new dimension-reduced space. Figure 6.13 shows the model performance versus the truncation size, or the number of eliminated

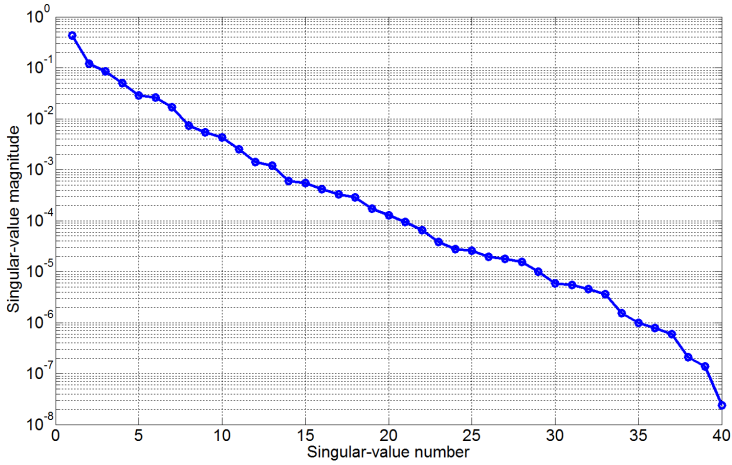


Figure 6.12: Singular values of input matrix in eq. (6.7).

singular values, used for matrix inversion. The RxBN model was the same as the one used in Fig. 6.10 with 40 terms. Hence, in the corresponding LS problem the rank of the input matrix was 40. Figure 6.13 shows that reducing the effective dimension of the LS problem by 30, we have only about 2 dB degradation in model performance.

6.6.3 Model pruning

In this section we investigate the effect of model complexity on the robustness to disturbance, and propose an approach to get the optimal solution for RxBN modeling.

In order to investigate the effect of disturbance on the model performance, we add artificial Gaussian noise in post-processing to our measured RxBN. Table 6.3 shows the MP model performance in terms of NRMSE against the disturbance to signal ratio (DSR). The DSR in this table is defined as the power ratio of the added disturbance inside the receive-band to the measured RxBN including both stochastic and deterministic part. The table shows that the model performance dramatically deteriorates as the DSR, namely the disturbance strength, increases.

As discussed earlier in Section V, the condition number is a good indication of model robustness to disturbance. This fact suggests a way to enhance the

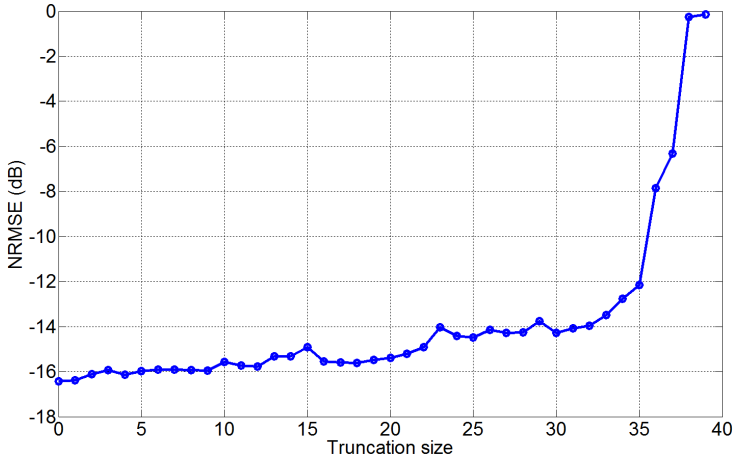


Figure 6.13: Performance of the RxBN model in which truncated SVD was used for matrix inversion. The performance is shown in terms of NRMSE versus the truncation size, i.e., the number of eliminated singular values. The RxBN model is the same as the one used in Fig. 6.10 with 40 terms.

Table 6.3: NRMSE of MP model versus disturbance-to-signal ratio (DSR)

DSR (dB)	-20	-10	-6	-2	2	6
NRMSE (dB)	-16.5	-16.1	-15.3	-13.7	-11.1	-7.8

robustness by removing some terms in the model.

We expect that by reducing the number of terms in eq. (6.10) the corresponding $\mathbf{X}$ in eq. (6.7) will have a smaller condition number, leading to enhanced robustness. However, different terms of eq. (6.10) do not generally have the same effect on the condition number. In our approach we eliminate each term of eq. (6.10), which is equal to eliminating the corresponding column of $\mathbf{X}$, and calculate the resulting $\kappa(\mathbf{X})$. This would enable us to find the terms of eq. (6.10) that are the most influential on the condition number. Fig. 6.14 illustrates the most influential elements in an MP model with $M = 5$ and $S_k = \{2, 4, 6, 7, 9, 11, 13, 15\}$ that was used in Section VI.A. The table is a representation of the terms present in the MP model. The columns represent the nonlinearity order and the rows represent the delay taps. This figure shows that the terms that are the most influential on the condition number are those corresponding to the higher nonlinearity orders and delay taps of one and two. For example the first two most influential ones are $|x(n-2)|^8 x(n-2)$

and $|x(n-2)|^{10}x(n-2)$. Note that the most influential term discussed above completely differs from the most significant term. The former has high impact on the condition number value, whereas the latter highly influences the modeling accuracy.

	$x\|x\|$	$x\|x\|^3$	$x\|x\|^5$	$x\|x\|^6$	$x\|x\|^8$	$x\|x\|^{10}$	$x\|x\|^{12}$	$x\|x\|^{14}$
x_n				28	18	13	15	16
x_{n-1}		32	29	20	4	3	5	10
x_{n-2}		30	19	8	1	2	7	11
x_{n-3}		27	21	12	6	9	14	17
x_{n-4}			31	26	23	22	24	25

Figure 6.14: Terms of MP model in eq. (6.10) that are the most influential on condition number. The model has $M = 5$ and $S_k = 2, 4, 6, 7, 9, 11, 13, 15$. The most influential term has the rank 1 [7].

The next step is to eliminate the first N terms that have the highest impact on the condition number value. This process is called pruning. By varying N we get the plot of Fig. 6.15. Note that Fig. 6.15 shows the condition number versus the number of eliminations, not the numbers shown in Fig. 6.14. The numbers in Fig. 6.14 show the importance orders of the terms in the MP model.

Figure 6.16 illustrates the effect of N, the number of eliminations, on modeling accuracy in two cases. The first one is the case that we model the measured RxBN without any added disturbance. The second one is the case that we add a Gaussian noise in post-processing with the power of 6 dB higher than the measured RxBN power. It can be expected that in the former case we lose accuracy by elimination. However, the latter case is interesting; here we gain better accuracy by elimination. It is clear that from a certain N onward the accuracy starts getting degraded, because in such a condition the degradation in modeling accuracy due to the elimination masks the reduction in sensitivity.

Based on Fig. 6.16 the optimum N can be chosen. We call the resulting model the robust MP (RMP). Depending on noise robustness requirements, one can use different values of N. Fig. 6.17 shows the NRMSE of the MP and robust MP versus the disturbance strength for three cases: $N = 0$, $N = 17$ and $N = 26$. The respective numbers of terms in the corresponding models are 40, 23 and 14. Table 6.4 is the numerical representation of Fig. 6.17. It is clearly seen that by pruning the MP model, we lose slight accuracy, but in return gain robustness

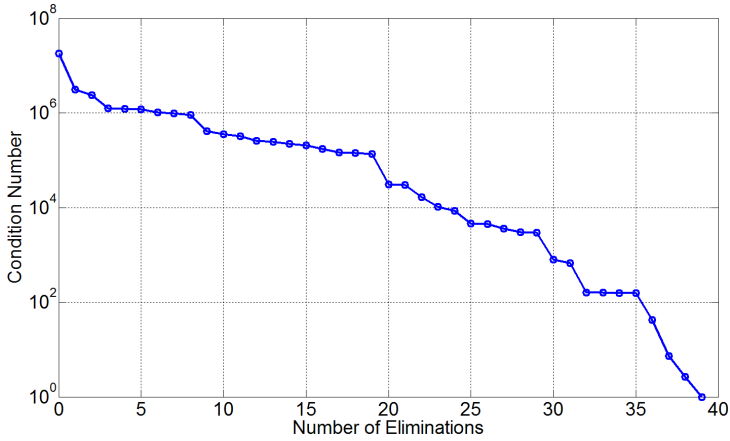


Figure 6.15: Condition number of $\mathbf{X}$ in eq. (6.7) for MP model in eq. (6.10) versus the number of eliminated columns that have the highest impacts on the condition number [7].

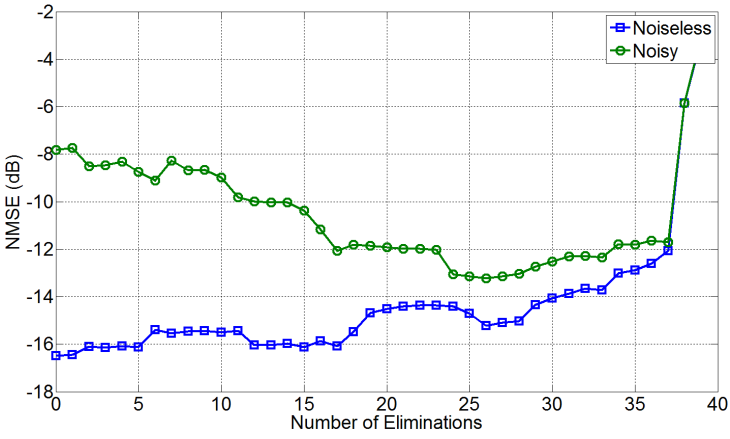


Figure 6.16: NRMSE versus the number of eliminated columns that have the highest impacts on the condition number. The noisy case is the case that a Gaussian noise with the power of 6 dB higher than the RxBN power is added [7].

to large disturbance. As the figure shows, the NRMSE remains better than -13 dB with disturbance-to signal ration (DSR) of -2 dB for $N = 0$, whereas this accuracy is still achieved with $DSR = 4$ dB for $N = 17$ and $DSR = 6$ dB for $N = 26$. The NRMSE degradation of the model is around 0.5 dB with $DSR = -10$ dB for $N = 0$, $DSR = -4$ dB for $N = 17$ and $DSR = 0$ dB for $N = 26$. With $DSR = +10$ dB we have a low accuracy of $NRMSE = -4.1$ dB for $N = 0$, yet $NRMSE = -9.1$ dB for $N = 17$ and $NRMSE = -11.2$ for $N = 26$. In other words, the modeling accuracy for $N = 26$ is degraded by only 4 dB with a disturbance 10 dB larger than the desired signal.

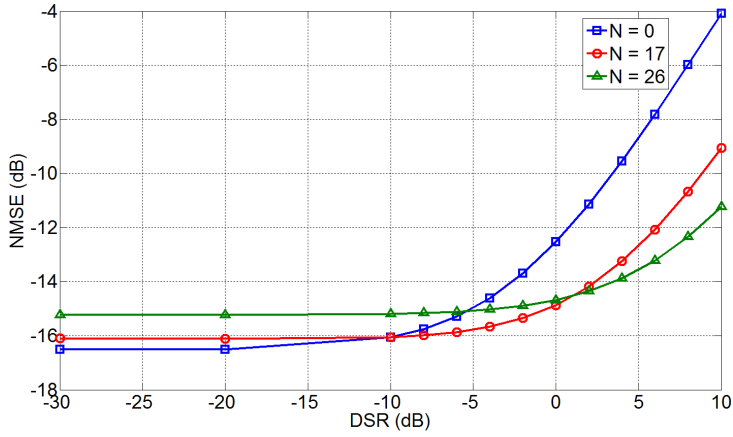


Figure 6.17: Modeling performance of MP and robust MP versus disturbance to signal ratio (DSR) with different numbers of elimination (N) for measured RxBN of PA output at 24 dBm [7].

Figures 6.18 and 6.19 show the results of the same investigation on PA2.

An alternative approach for finding the right terms to eliminate in the model is by eliminating each term, estimating the model coefficients and calculating the resulting modeling error. At first glance, this method seems the most reliable way of choosing the right terms for elimination. However, this method is very time-consuming and computationally intensive. Moreover, it is not guaranteed that the best elimination at one step leads to the best final solution; it might happen that an elimination is not the best at a certain step, but leads to the best solution after further elimination. This means that for finding the global maximum of the problem, we may require an exhaustive search which is extremely inefficient in time and computation. The proposed approach is fastest and yields satisfactory results for RxBN modeling. In our approach we

Table 6.4: NRMSE of MP and robust MP model versus DSR with different N for measured RxBN of PA output at 24 dBm

DSR (dB)	NRMSE (dB)		
	$N = 0$	$N = 17$	$N = 26$
-30	-16.5	-16.1	-15.2
-20	-16.5	-16.1	-15.2
-10	-16.1	-16.1	-15.2
-8	-15.8	-16.0	-15.2
-6	-15.3	-15.9	-15.1
-4	-14.6	-15.7	-15.0
-2	-13.7	-15.4	-14.9
0	-12.5	-14.9	-14.7
2	-11.1	-14.2	-14.4
4	-9.5	-13.3	-13.9
6	-7.8	-12.1	-13.2
8	-6.0	-10.7	-12.4
10	-4.1	-9.1	-11.2

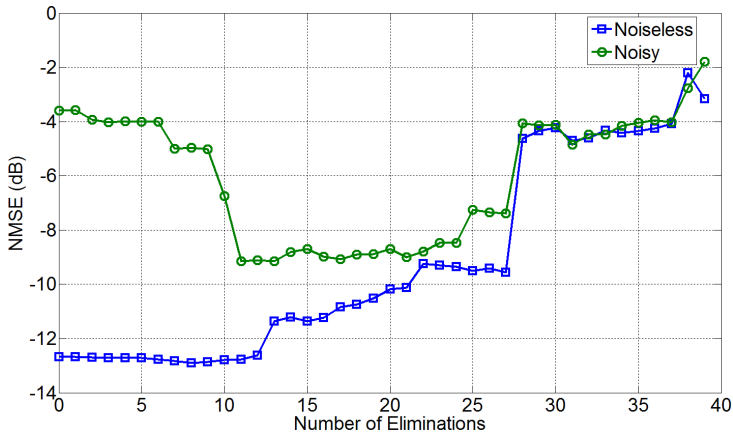


Figure 6.18: NRMSE versus the number of eliminated columns that have the highest impacts on the condition number. The noisy case is the case that a Gaussian noise with the a power of 6 dB higher than the RxBN power is added [7].

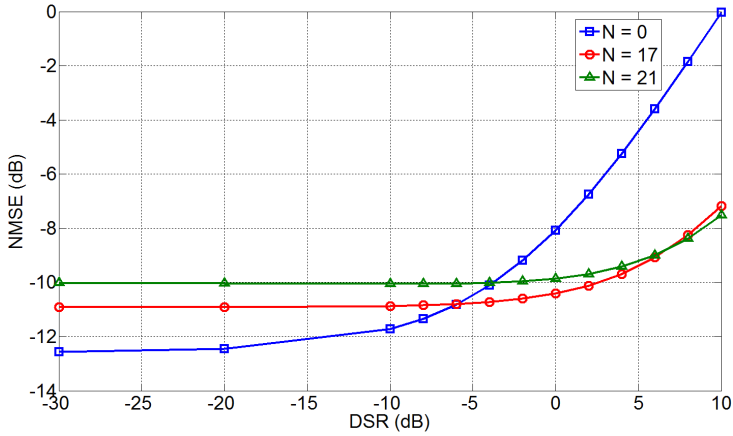


Figure 6.19: Modeling performance of MP and robust MP versus disturbance to signal ratio (DSR) with different numbers of elimination (N) for measured RxBN of PA output at 24 dBm [7].

require only calculating the condition number after each elimination without the need for estimation and error calculation. Also we calculate the condition numbers at only one step, and choose all the terms to be eliminated based on that single-step calculation.

The robust MP identification procedure has an additional step for finding the most influential term. This addition can be insignificant, since it is not required to be run frequently. We assume that the model architecture does not change frequently over time for a given average output power level, whereas the model coefficient values may drift over time more often, for example due to environmental changes. By model architecture we mean the set of terms that are used in the RxBN model, and the model coefficients are the weights of these terms. Possible change in model architecture at each power level is assumed to be much slower than drift in coefficient values. Therefore, we can benefit from the simpler coefficient estimation of the robust MP.

In summary, the robust MP has two major advantages over MP: 1) its performance is robust to relatively large noise. In fact, it compromises the modeling accuracy of RxBN with low stochastic noise for the modeling accuracy of RxBN with high stochastic noise. 2) the model size is smaller than that of full MP, resulting in lower computational complexity.

6.6.4 Benchmarking of RxBN model

Here, by benchmarking we mean comparing our model, as much as possible, with the approaches available in the literature. Since the topic of RxBN modeling is new, and its potential effectiveness for performance improvement in emerging and future telecommunication technologies has been recently felt, there are not enough relevant published results to have a plausible benchmarking. To the best knowledge of this PhD thesis' author, there are published measurement results for RxBN modeling from only one research group which published their works in two conference papers [43, 45]. Of these two papers, only the one reported in [45] has conditions comparable with the condition of our measurements in terms of signal bandwidth. However, they did not mention duplex spacing. Apart from the lack of mentioning the measurement conditions, they reported -18 dB of accuracy. It should be noted that the modeling accuracy is dependent on waveform dynamics: using a more complex modulation scheme leads to lower accuracy. We used 4-QAM modulation. The mentioned paper said that they used a High Speed Uplink Packet Access (HSUPA) waveform. This type of waveform can have binary phase-shift keying (BPSK) or quadrature phase-shift keying (QPSK) modulation. In summary, since the measurement conditions of [45], especially the modulation type and duplex spacing, are not fully mentioned, we cannot have a fair comparison of our modeling performance with their results. Furthermore, the paper did not mention the detail of the model that they used, which is the lack of another key criterion for comparison. However, as mentioned before, in this section we have provided a comprehensive strategy for characterization and efficient modeling of deterministic RxBN, which are totally absent in [45].

6.7 RxBN cancellation

As explained in Section 6.1, RxBN can degrade the receiver performance, and should be canceled. Since in some cases, such as Band 13 of the LTE standard discussed in Section 6.1, RxBN cannot be easily mitigated in circuit design, there are increasing efforts to mitigate it at system level using different signal processing techniques. This sectione presents and discusses two approaches for RxBN cancellation at system level.

6.7.1 Feedforward RxBN Cancellation

There are increasing efforts to mitigate the RxBN in transceivers using different techniques, most of which are based on feed-forward cancellation (FFC) [43, 44, 46, 119]. These techniques are based on copying the generated RxBN in the RF path and subtracting it from the received signal, after gain and phase compensation, either in the analog or the digital domain. Figure 6.20 shows an FFC scheme in which cancellation is done in the digital domain by using digital signal processing (DSP) techniques [45].

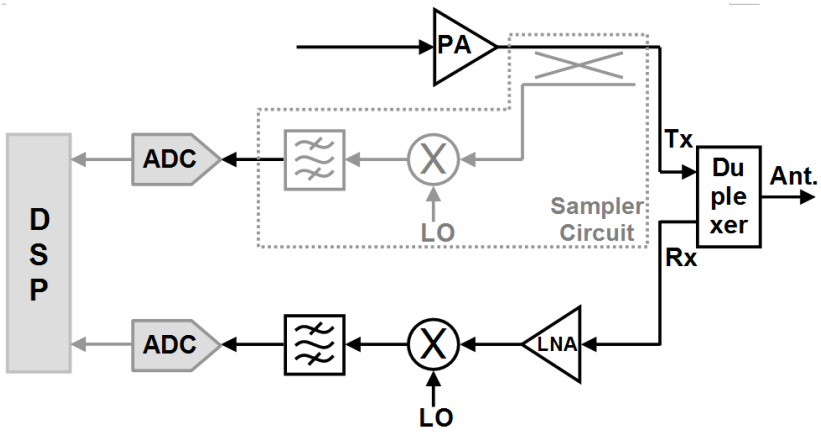


Figure 6.20: Diagram of feedforward RxBN cancellation approach which utilizes RxBN modeling for cancellation [7].

As shown in this figure, a sampling circuit gets a sample of the PA output for RxBN measurement, using a coupler followed by a receiver circuit which downconverts the spectral contents of the PA output in the Rx band for subsequent analog-to-digital conversion (ADC). The digitized data is then subtracted in DSP from the digitized received signal coming from the main receiver. There are two approaches for subtraction: blind subtraction and estimated subtraction. In blind subtraction, the sampled output is directly subtracted from the received signal after compensation of gain and phase for the electrical difference between the sampling circuit path and the main receiver path. In estimated subtraction, the sampled signal is used to build a model for the estimation of the RxBN in the main receiver [42, 45]. The estimated RxBN is then subtracted from the received signal in DSP. Using the estimated RxBN is beneficial since it relaxes the specification requirements of the sampler circuit (depicted in Fig. 6.20), because of the robustness to unwanted noise. For

estimated FFC, the modeling approach already proposed in this chapter can be used.

6.7.2 Transmitter Signal Injection

Another approach for RxBN cancellation could be injection of a signal along with the input in the transmitter in a way that it cancels out the RxBN caused by the Tx nonlinearity. This technique is similar to DPD, with the difference that in this case the region of cancellation is the Rx band, not the frequency range around the Tx band. The technique is referred to as transmitter signal injection (TxSI), and is illustrated in Fig. 6.21. As shown in the figure, injection in the Rx band at the input can diminish the RxBN caused by the PA nonlinearity and excited by a Tx-band input signal. The injected signal requiers to have appropriate amplitude and phase for cancellation.

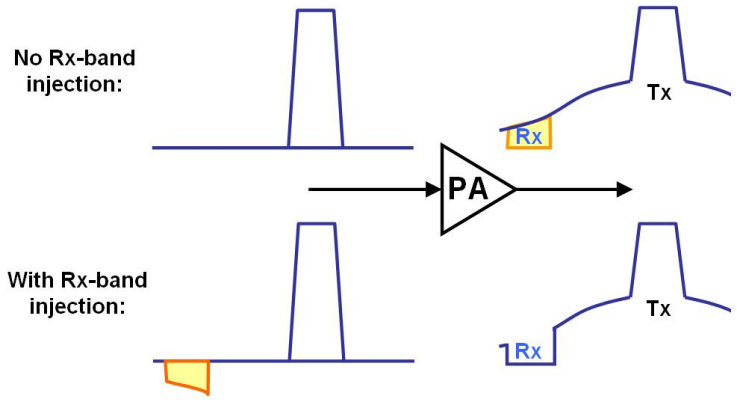


Figure 6.21: TXSI approach for RxBN cancellation: the upper row is a typical case, without any cancellation scheme, with RxBN present at the output. The lower row is after signal injection in the Rx band in such a way that the output RxBN is diminished.

Obviously, injection in the TXSI approach is applied before the generation of RxBn. Therefore, for a correct injection one requires to know what will be generated as RxBN at the ooutput. In other words, one should estimate the RxBN in this approach. Hence, the TXSI approach is based on RxBN modeling, as illustrated in Fig. 6.22. First, an auxiliary receiver (Aux. Rx) gets a sample of the output RxBN. After downconcersion to the baseband digital data, it is used for the identification of the RxBN model in DSP. The built RxBN model

then is used for generattion of the appropriate baseband Rx-band injected signal which is generated by an auxiliary digital-to-analog converter (Aux. DAC), and is added to the Tx-band PA input signal.

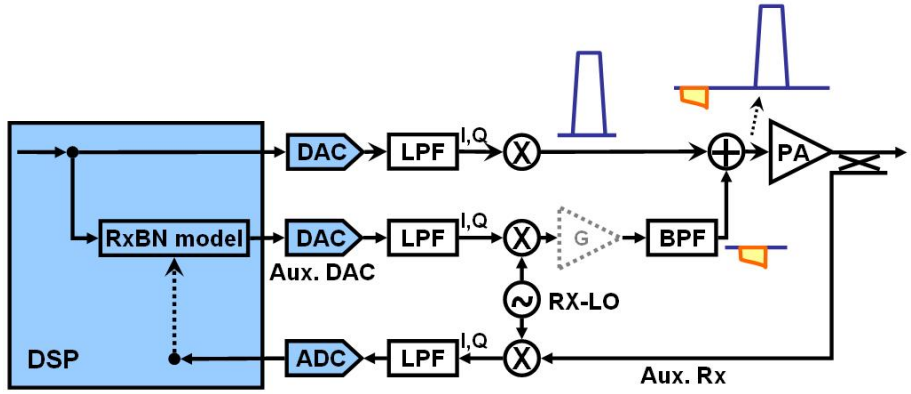


Figure 6.22: Diagram of RxBN cancellation by using TXSI approach presented in Fig. 6.21.

A consideration should be made in the TXSI procedure. The injected signal should be compensated for gain and phase distortion caused by different paths. Figure 6.23 shows different building blocks of the presented TXSI technique, composed of an auxiliary receiver (Aux. Rx), a DSP block and a signal injector (SI) block. Each building block has some linear distortion effect (in both gain and phase) on the signal. Hence, all these effects should be taken into account for a successful cancellation. To do so, an equalization step is added to the procedure. In this step, a training sequence in the Rx band is sent to the system, while the system is excited by a Tx signal to keep the correct operating point. The training Rx-band signal should be small enough so that it does not change the operating point. This *calibration* process also compensates for the delay caused by the block. The delay is important since the injected Rx-band signal should be applied at the correct time. Otherwise, it cannot cancel the generated RxBN perfectly.

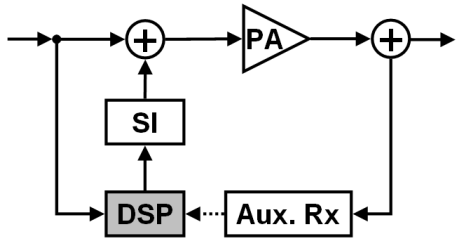


Figure 6.23: Building blocks of TXSI technique presented in Fig. 6.22.

6.7.3 Case study: TXSI with multisine excitation

Employed TXSI architecture

The proposed TXSI technique is tested with 11-tone multisine excitation and 10 MHz bandwidth. We assume band 13 frequency settings described in Section 6.1 for this test. Instead of using an auxiliary DAC for injection, we use the same VSG that is used for the generation of Tx signal (the 11-tone excitation in this case). This source has enough bandwidth for simultaneous generation of both the Tx and injected Rx signals. However, it should be noted that, as can be seen in Fig. 6.2, a filter is inserted after the VSG to suppress signals outside the Tx band. Therefore, the injected Rx-band signal is attenuated as well by this filter. This is not a problem as long as we are injecting small signals. A larger Rx-band signal can be generated at the VSG such that after the filter attenuation the resulting signal is the desired one. The filter response and its linear distortion effect is also compensated in the calibration process described above.

Equalization

The training sequence used in this test is illustrated in Fig. 6.24. It is composed of the main Tx-band signal and a filtered Gaussian noise centered at Rx-band. Basically, the Rx-band training signal should be small enough so that it does not change the operating point of the PA. However, the spectrum shown in Fig. 6.24 has a relatively large Rx-band signal. The point in this case is that the signal in this figure was the VSG output, and it is filtered before exciting the PA. Therefore, the Rx-band signal after attenuation by the filter is small enough. The resulting signal then excites the PA, and will be further affected by linear distortions along the whole path by different components. The captured

Rx-band signal at the VSA is then compared with the original one at the VSG to find a model, called *equalizer*, to compensate the total linear distortion caused by the entire path. The equalizer is identified by indirect learning, explained

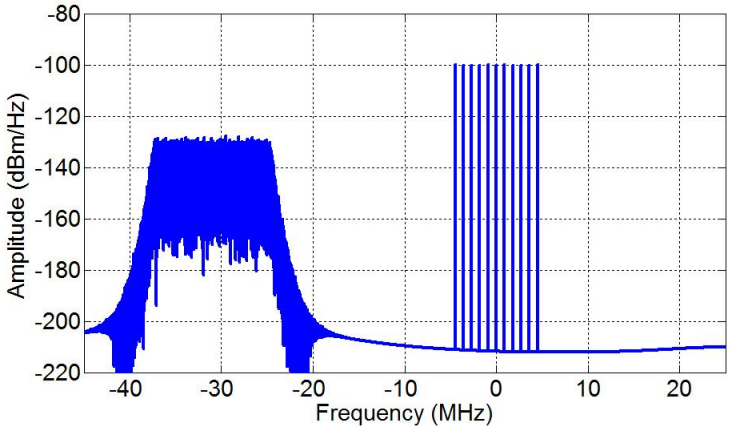


Figure 6.24: Training sequence used for calibration.

in Section 2.5. Contrary to the nonlinear models considered in Section 2.5, in this case the model is linear, since it is aimed at the compensation od linear distortion. The equalizer model is of the form

$$v_n = \sum_{m=0}^{M} a_m u_{n-m} \tag{6.11}$$

where M is the memory depth. Equation (6.11) is a simple convolution with limited memory depth. The variables u_{n-m} and v_n are the equalizer input at the time instance $n-m$, and the equalizer output at the time instance n, respectively, as shown in Fig. 6.25. As can be seen in the figure, the equalizer, denoted by "Eq.", is implemented in the DSP part of the chain. In essence, the equalizer block is the one added to the TXSI path to compensate its linear distortion. In eq. (6.11), the weights a_m can be identified by a least square method.

Using the waveform illustrated in Fig. 6.24, we have been able to get the equalization error of -23 dB in our application.

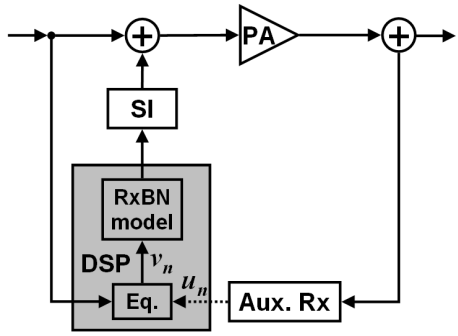


Figure 6.25: Block diagram of the TXSI architecture with equalizer.

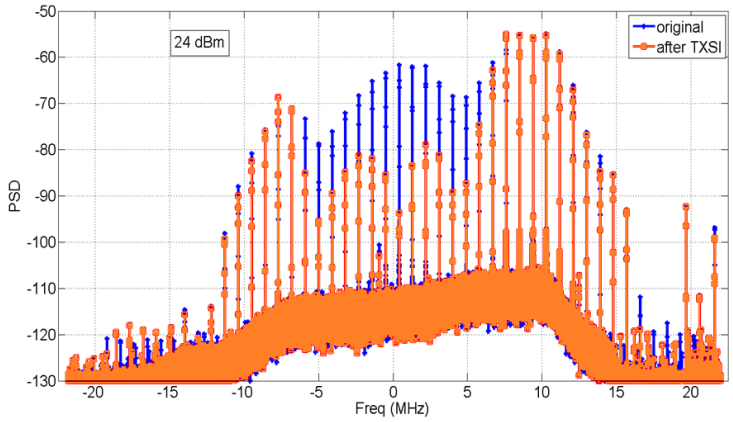
Table 6.5: Modeling accuracy and RxBN cancellation by using TXSI technique at different PA output power levels with 11-tone excitation.

P_{out} (dBm)	23	24	25	26
Model error (dB)	-16.1	-17.3	-18.3	-19.5
Tone power reduction (dB)	16.3	17	17.4	19.2
Total power reduction (dB)	13.8	14.5	15.7	17.0

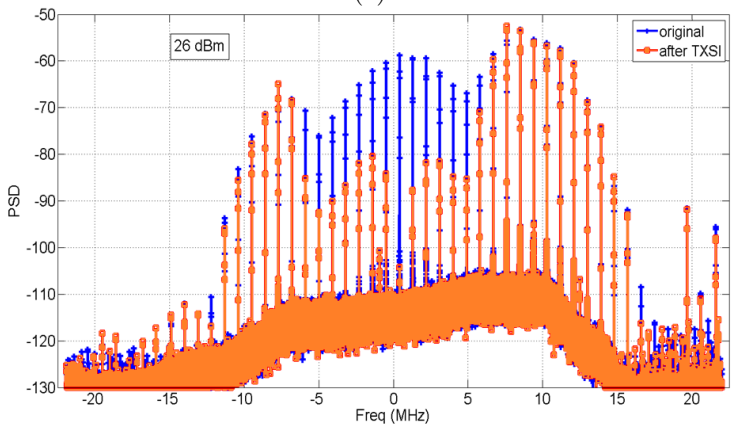
RxBN cancellation results

After calibrating the path by the equalizer, the estimated Rx-band signal is injected to the input of Fig. 6.25. The RxBN output is cancelled to some extent at the PA output, as shown in Fig. 6.26. In this figure, the TXSI cancellation is applied at two different power levels of 24 dBm and 26 dBm. Table 6.5 shows the amount of cancellation at these two and two more PA output power levels. In this table, the accuracy of the model used for TXSI technique at each power level is indicated. The second row of data shows the cancellation amount at the tone locations, while the third row shows the cancellation amount with respect to the total RxBN which includes the stochastic noise. Obviously, the total power cancellation is smaller than the tone power cancellation, since in the former case the effect of stochastic noise, which cannot be mitigated, is taken into account.

As can be seen in Fig. 6.26 and table 6.5 there is still room for better cancellation, and the RxBN tones in the figures are not cancelled perfectly. There are two main reasons for this limited cancellation. First, the accuracy of the model used for TXSI cancellation is limited, as can be seen in table 6.5. This limitation



(a)



(b)

Figure 6.26: RxBN cancellation by using TXSI technique at the PA output power level of a) 24 dBm, and b) 26 dBm.

means that the cancellation also cannot be better than the accuracy of the model. Second, time alignment between the Rx-band injection and the Tx-band input signal is very critical. The Rx-band signal should be injected at the correct time to properly cancel the output RxBN. If there is some misalignment between the two signals, the cancellation will be degraded. Therefore, in order to improve the cancellation performance, the modeling accuracy and the time time alignment should be improved.

6.7.4 Comparison and discussion

The two proposed techniques, namely digital FFC and TXSI, have some basic differences which are addressed in the following.

- The digital FFC is essentially a post-processing technique, while the TXSI is applied with the input of the system.
- Duplexers have ususally a complicated linear frequency response. Hence, the duplexer presence in a cancellation approach exacerbates the difficulty of accurate performance. The digital FFC includes the duplexer in the chain, while the TXSI is performed before the duplexer. However, a filter may be needed in the auxiliary receiver path of the TXSI to reduce the high power level of the Tx-band signal, otherwise the large Tx-band signal can easily saturate the auxiliary receiver. This additional filter can also exacerbate the complexity, but can have a smaller complication effect.
- stochastic RxBN can be cancelled in an appropriate digital FFC approach (blind cancellation), but cannot be cancelled in the TXSI, since this technique is based on the deterministic RxBN modeling.
- TXSI requires an additioanl DAC (auxiliary DAC), which means an additional hardware and power consumption.

6.8 Conclusions

Receive-band noise (RxBN) in receivers caused by non-idealities of transmitters have been discussed in this chapter. There are two types of RxBN: stochastic and deterministic. These two types have been investigated, and some methods to distinguish them from each other are proposed. A technique based on the offset multisine concept is used to identify the important nonlinearity orders

that contribute to the deterministic RxBN. It has been shown that modeling the deterministic part can be used for RxBN cancellation schemes. Hence this modeling has been investigated in this chapter. A detailed approach for deterministic RxBN modeling is presented and the important issues to perform adequate modeling are explained. The model is a modified memory polynomial model which includes even-order terms. A novel strategy for efficient model pruning targeting at robustness to disturbance is presented. All the presented techniques for different parts of modeling procedure are experimentally validated, and the proposed model is successfully tested on different amplifiers.

RxBN cancellation is the ultimate goal of this chapter's discussion. Two different approaches for RxBN cancellation have been explained. One approach mitigates the RxBN at the transmitter, while the other one does it at the receiver. The two approaches are compared and discussed in this chapter.

Chapter 7

Conclusions

7.1 Research summary

Non-ideal effects can be detrimental to the performance of the RF circuits and systems. If they are not considered in system design, the system may not work even sub-optimally. The objective in this PhD thesis was to characterize, model and mitigate non-ideal effects in microwave circuits and systems. The related four-year research conducted by the author in this direction is summarized and reported in the chapters of this PhD thesis as follows.

In Chapter 3, modeling and linearization around carrier frequency was investigated. There are two approaches for mitigating non-ideal effects: mitigation in circuit-level design, and mitigation in system-level design. The latter was the subject of that chapter, as well as the rest of this PhD thesis. For system-level mitigation, modeling the non-ideal effects may be required. Different modeling approaches were presented in Chapter 3. More specifically, the chapter contributed to resolving the classic problem of modeling and linearization with the following:

- A novel formulation of S-functions for modeling intermodulation effects around the carrier frequency in the multitone excitation case was proposed and validated in simulation. The proposed formulation takes into account only the large tones and their intermodulation products around the carrier frequency. In other words, unlike the classic S-functions formulation it does

not include harmonics, and assumes that they are suppressed sufficiently. The method was tested on a mixer, seen as a two-port component with internal LO, in simulation. The results validated the performance of the approach. Then linearization is performed based on this formulation. The intermodulation products were successfully suppressed with this S-functions-based linearization technique. This work was presented in the 6^{th} European Microwave Integrated Circuits Conference (EuMIC), 2011 [1]. The PhD thesis' author contributed to the work with proposing the new S-functions formulation for modeling the intermodulation products around the fundamental frequency. Also the work of [1] was supervised by the PhD thesis' author.

- A novel Volterra-based model, referred to as cross-memory polynomial (CMP), was proposed and validated. The model partly includes cross-terms, which can improve the model performance. The model performance was experimentally validated, and its performance was shown to be comparable with the established memory polynomial (MP) method. This work was presented in the 7^{th} European Microwave Integrated Circuits Conference (EuMIC), 2012 [2].

- An FPGA-based digital predistortion (DPD) architecture is presented. The DPD system were successfully validated in experiment, employing both the CMP model and the MP model. The performance of the CMP-based DPD was again proved to be comparable with that of the MP-based DPD. This work was presented in International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits (INMMIC), 2014 [3]. The work was supervised by the PhD thesis' author.

- An efficient approach for modeling an array of amplifiers was proposed. The approach uses a common model, to model the common behavior of the array amplifiers, along with additional models for each individual amplifier to model their specific different behavior. The performance of the approach is validated with experiments. This research was conducted in collaboration with the National University of Ireland, maynooth, and the results were presented in IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR), 2013 [4].

The first step in system design, is to have an appropriate understanding of non-ideal effects, namely characterizing them. A novel method for characterization of some of those effects was presented and validated in Chapter 4. In this method, the tone locations of offset multisine excitation are slightly offset in a way that the intermodulation products of the tones lie at separate unique locations. In this way, we are able to characterize the behavior of the nonlinearity in more

detail and with deeper understanding. Some interesting applications of this method, particularly memory effects characterization and assessment of model strength in estimating memory effects, are demonstrated with experimental results. This research was conducted in collaboration with the University of California, San Diego, USA, and the results were published in the IEEE Transactions on Microwave Theory and Techniques, March 2014 [5].

In addition to the above-mentioned classic problem of modeling and characterization around the carrier frequency, two more applications of modeling and signal processing in RF systems are presented in this PhD thesis. They are: below-noise characterization of data-converters by using dithering, and receive-band noise (RxBN) cancellation in RF transceivers.

Chapter 5 explained dithering techniques for reduction of quantization noise in data converters. Then a novel dithering technique was presented for characterization purpose. The technique is based on adding a small dithering signal to the input, and performing some post-processing procedure. After this procedure, the obtained signal will have better spectral purity with reduced quantization noise and non-ideal effects caused by dynamic nonlinearity, and probably memory, of the data converter. The technique was successfully tested with a large-bandwidth signal, and considerable spectral purity improvements was obtained.This research was conducted in collaboration with the National Institute of Standards and Technology (NIST), USA, and the results were published in the IEEE Transactions on Microwave Theory and Techniques, November 2013 [6].

In the last research chapter of this PhD thesis, namely Chapter 6, a topic at the cutting edge of the system-level improvement of RF front-ends was covered. The topic is RxBN cancellation in RF transceivers, and is increasingly getting more and more attention. RxBN is the transmitter emission in the receive band caused by transmitter nonlinearities. The emission can be so high that it degrades the receiver performance. This effect is more pronounced in recent and future telecommunications, where the bandwidth can be large and the relative duplex spacing can be small. The chapter proposed methods for characterization of receive-band noise. The methods enabled us to distinguish deterministic from stochastic RxBN, and to identify nonlinearity orders that have considerable contribution to the deterministic RxBN. A comprehensive approach for receive-band noise modeling was also proposed, which has acceptable performance and is robust to noise. Robustness to noise can be a determining factor in real applications, where the feedback sampling circuit has noise power comparable with the RxBn power. The proposed methods and approaches were all validated by experimental results. Moreover, methods for cancellation

of RxBn RF transceivers were presented in that chapter. One of the proposed cancellation methods was also implemented and tested experimentally, and showed acceptable performance with multisine excitation. This research was conducted in collaboration with the University of California, San Diego, USA, and the results were submitted to the IEEE Transactions on Microwave Theory and Techniques [7].

7.2 Future works

Possible directions for future research related to different parts of this PhD thesis are listed in the following. They are divided in four categories, corresponding to the four research chapters of this PhD thesis, namely Chapters 3, 4, 5 and 6.

7.2.1 Modeling and linearization around main band

S-functions were used in Chapter 3 for modeling and compensation of intermodulation products around main band. The formulation was adopted for multitone waveforms with equal large-tone amplitudes. The approach can be generalized to relax the amplitude equality restriction. Furthermore, the dependency of intermodulation on the frequency spacing can also be included in the formulation. In this way, memory effects are taken into account. The validation of the proposed approach was done in simulation. The next step is to validate it in experiment.

An FPGA-based DPD architecture was proposed in Chapter 3. This architecture did not include the DPD block yet, and this block was implemented in software. The next step is to include the whole DPD system in FPGA, so that we have a full-blown stand-alone FPGA DPD system.

7.2.2 Offset multisine

The proposed novel method of offset multisine is expected to find many interesting applications. The frequency locations of offset multisines were found by a search procedure. An interesting future work could be finding an analytical way to find the locations. In this way, finding the locations in the

case of increasing the number of tones, or taking into account the higher-order intermodulation products will be easier.

A very interesting potential application of offset multisine is in the identification of frequency-domain Volterra kernels. Since offset multisine excitation can distinguish different intermodulation products, it is a very well-suited tool for frequency-domain approaches. An extensive research can be done to employ offset multisine for the identification of frequency domain Volterra kernels, and to find ways to improve their accuracy. This research can be a breakthrough in Volterra-based modeling.

7.2.3 Time-domain ensemble average dithering

The dithering technique proposed in Chapter 5 was tested with a wideband signal, with the bandwidth of 160 MHz. This wide band led to more pronounced non-ideal effects which limited the improvement by the proposed technique. More enhanced performance is expected to be achieved by using smaller bandwidth, which more complies with the current telecommunication standards. Future works in this field could be:

- use of smaller bandwidth and investigating the maximum achievable improvement by the proposed dithering technique,
- investigation of the effect of dithering on different non-ideal effects separately, in other words, finding a way to be able to separately examine the effect of dithering on different non-ideal effects, particularly quantization noise and memory (hysteresis) effects,
- finding a better dithering signal to get larger improvement.

7.2.4 Receive-band noise modeling and compensation

As mentioned earlier in Chapter 6, RxBN cancellation is immature, and there is plenty of room to develop this technique. Here are some suggestions for future works:

- Deterministic and stochastic noise behavior against varying different parameters of the RF circuit, and their values compared with each other

can be further investigated. The varying parameters could be, but are not limited to, the power level, the DC current values of different blocks, modulation complexity, duplex-spacing-to-bandwidth ratio, etc.. Better characterization and understanding of stochastic and deterministic RxBN can lead to more efficient cancellation techniques.

- An interesting possibility is to combine the FFC and TXSI approaches to have a hybrid technique. This combination could enhance the cancellation performance.
- The model for estimating the deterministic part of RxBN is still immature. In different occasions in this chapter we have seen that there is still room for improvement, since the residual error is still above the stochastic RxBN level. Efforts can be made to improve the RxBN model performance.
- Time alignment has special importance in some approaches, particularly in the TXSI. Increasing the accuracy of time alignment will enhance the cancellation performance.
- Improving the accuracy of the equalization or calibration process also can considerably enhance the cancellation performance. As mentioned earlier, this procedure compensates the linear distortion caused by different blocks along the path.
- The TXSI approach is tested with 11-tone excitation. A step forward is using more realistic modulation, such as LTE waveforms and investigating the effect of modulation type on the cancellation performance. More generally, the effect of modulation on the accuracy of different cancellation techniques can be investigated.

As a more general and longer-term direction for future research, most of the techniques and approaches proposed in this PhD thesis should be adapted to emerging requirements for future RF systems. Those requirements often concern lower power consumption and larger bandwidth. RF systems with low power consumption are highly desired in order to achieve low cost, long power supply life and low CO_2 emission. Exponential increase of data rates demand for large total bandwidth. This bandwidth can be provided as a merged wide band, or an aggregation of multiple bands which are not necessarily adjacent to each other. Therefore, the approaches proposed in this PhD thesis will be very useful in future applications, if they are adapted to be energy-efficient and to include large bandwidth or multiple bands. In the sense of low power consumption, low-complexity, yet high-performance, schemes should be developed. A valuable research could be investigation and estimation of real power consumption of the DSP blocks used for DSP-based compensation. Then based on such information,

energy-efficient approaches can be sought. In the sense of large bandwidth, wide-band compensation with band-limited hardware is a highly demanded research. Furthermore, compensation in presence of multiple bands is still immature and requires intensive efforts.

Appendix A

Delay calculation in two-tone excitation

As mentioned in Section 3.2.1, the S-functions formulation derived for two-tone excitation assumes zero phases for the two large tones. Therefore, since the measured data does not necessarily fulfill this assumption, we have to calculate the time delay required to be added to the captured signals in post-processing, in order to get zero phases. This appendix proposes a method of delay calculation.

Assume that the frequencies of the two large tones are f_1 and f_2, their amplitudes are one, and their initial phases equal zero. The tones can then be represented by:

$$x_1(t) = \sin(2\pi f_1 t) \tag{A.1}$$

$$x_2(t) = \sin(2\pi f_2 t). \tag{A.2}$$

After passing the time t_0, the phases of the tones differ from each other, according to:

$$\phi_1 = 2\pi f_1 t_0 - 2n\pi \tag{A.3}$$

$$\phi_2 = 2\pi f_2 t_0 - 2m\pi, \tag{A.4}$$

where n and m are integer numbers equal to or greater than zero, and $0 \leq \phi_1, \phi_2 < 2\pi$.

Because the waveforms are captured at an arbitrary time in experiments, the phases of the tones are generally nonzero, and we require to compensate the delay causing such phases. In other words, we are looking for the minimum delay τ such that:

$$2\pi f_1 \tau = 2n\pi + \phi_1 \tag{A.5}$$

$$2\pi f_2 \tau = 2m\pi + \phi_2, \tag{A.6}$$

with n and m as small as possible. To find the minimum values for these parameters, we should know that starting from the initial zero phases, in what period T the two large tones simultaneously return to their initial phase condition. Figure A.1a and b show the amplitude and phase of two tones with $f_1 = 3$ Hz and $f_2 = 4$ Hz, and Fig. A.1c and d show the same plots for $f_1 = 3$ Hz and $f_2 = 5$ Hz. Along with the phase of the tones, the phase difference between the tones is also plotted in Fig. A.1b and d. Clearly, the phase difference is periodic with the period of $T_p = |f_1 - f_2|^{-1}$. However, it does not necessarily mean that the period T that we are looking for has the same value. This fact is seen in Fig A.1d where $T = 2T_p$, while in Fig A.1b we have $T = T_p$. Generally, T is obtained by:

$$T = \frac{1}{\mathrm{GCD}(f_1, f_2)}, \tag{A.7}$$

where GCD(.) is an operator denoting the greatest common divisor (GCD). Note that $\mathrm{GCD}(f_1, f_2) \leq |f_1 - f_2|$, hence $T \geq T_p$, or more accurately:

$$T = pT_p \tag{A.8}$$

$$T_p = |f_1 - f_2|^{-1} \tag{A.9}$$

$$p = \frac{|f_1 - f_2|}{\mathrm{GCD}(f_1, f_2)} \quad , \quad p \in \mathbb{Z} \quad , \quad p > 0, \tag{A.10}$$

with p being an integer number greater than zero.

Considering the above-mentioned discussion, we now find the minimum required τ. Subtracting eq. (A.5) from eq. (A.6), we get:

$$2\pi(f_1 - f_2)\tau = 2l\pi + (\phi_1 - \phi_2) \tag{A.11}$$

$$l = n - m. \tag{A.12}$$

If $\mathrm{GCD}(f_1, f_2) = |f_1 - f_2|$, p in eq. (A.8) equals one. In other words, the period of $\phi_1 - \phi_2$, which is T_p, is equal to T. In this case, since we want to find the

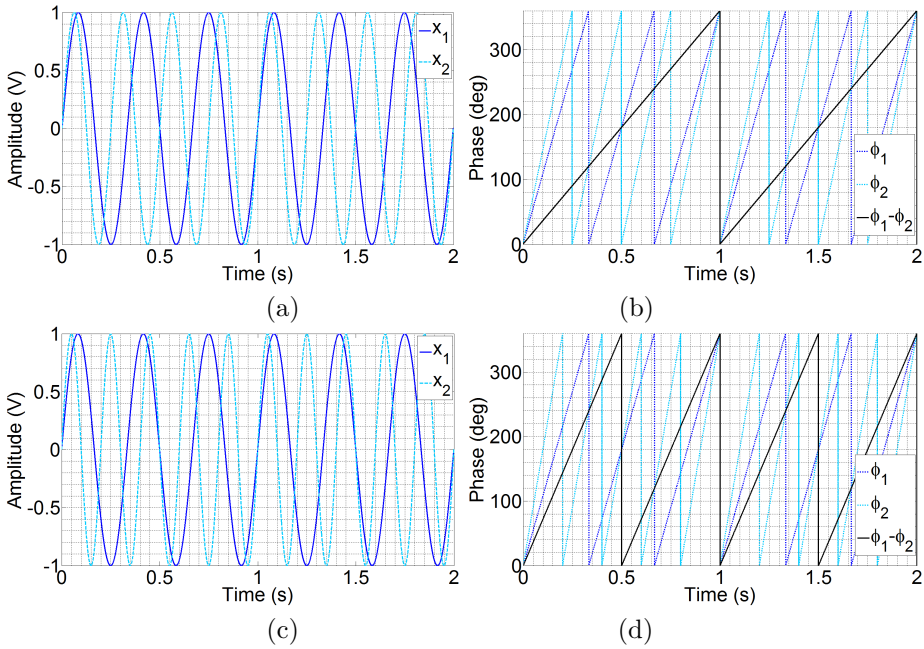


Figure A.1: Time-domain illustration of two tones: a) amplitude illustration with $f_1 = 3$ MHz, $f_2 = 4$ MHz, b) phase illustration with $f_1 = 3$ MHz, $f_2 = 4$ MHz, c) amplitude illustration with $f_1 = 3$ MHz, $f_2 = 5$ MHz, and d) phase illustration with $f_1 = 3$ MHz, $f_2 = 5$ MHz.

minimum τ, we should set $l = 0$. Therefore, if $\mathrm{GCD}(f_1, f_2) = |f_1 - f_2|$:

$$\tau = \frac{\phi_1 - \phi_2}{2\pi(f_1 - f_2)}. \tag{A.13}$$

In a more general case, if $\mathrm{GCD}(f_1, f_2) \neq |f_1 - f_2|$, there are more than one period of $\phi_1 - \phi_2$ in one period of T. Hence, knowing the value of only $\Delta\phi = \phi_1 - \phi_2$ does not provide enough information for calculating τ, since we do not know in which period of T_p in the period T the difference $\phi_1 - \phi_2$ has occurred (we do not know the value of l). Thus, first we have to find the period of occurrence. To find the l value, we need to consider the absolute values of ϕ_1 and ϕ_2.

Assume that the tones are captured at the time instant t_s, with the respective phase values of ϕ_1 and ϕ_2, as shown in Fig. A.2. In this figure, $T = 1$, $p = 3$ and $T_p = 1/3$. There are three points (A, B and C) in one period of T, at which the phase difference $\Delta\phi = \phi_1 - \phi_2$ equals zero. However, only at the last point

C the phases of the tones have simulataneously returned to their initial zero values, as can be seen in the figure. Therefore, the required delay τ is equal to t_s, not $t_s - t_B$ or $t_s - t_A$, where t_A and t_B are the time of point A and B, respectively.

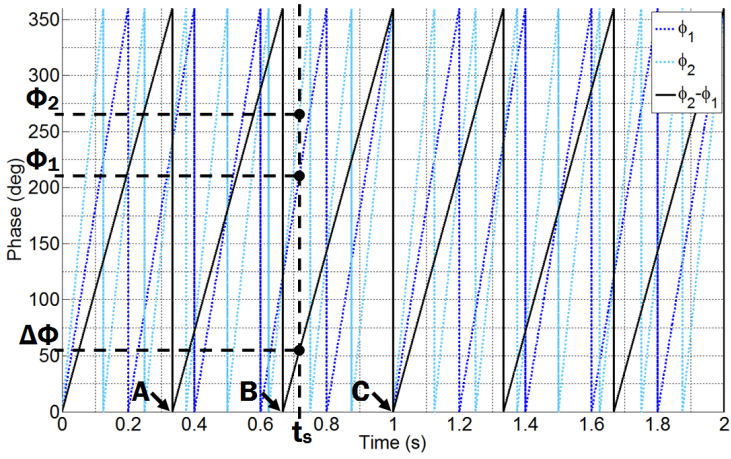


Figure A.2: Time-domain illustration of the pahses of two tones with $f_1 = 5$ MHz and $f_2 = 7$ MHz, along with their difference.

To find the value of t_s we follow two steps. First, the time distance τ_0 to the closest point with zero value of $\phi_1 - \phi_2$ (point B in the case of Fig. A.2 with $\tau_0 = t_s - t_B$) is calculated. Then the number l of T_p periods between that point and the initial zero phase point will be found. For example, $l = 2$ in the case of Fig. A.2. For the first step we have:

$$\tau_0 = \frac{\phi_1 - \phi_2}{2\pi(f_1 - f_2)} = \frac{\phi_1 - \phi_2}{2\pi} T_p. \tag{A.14}$$

The absolute phase at the point of closest zero $\phi_1 - \phi_2$ is:

$$\phi_0 = \phi_1 - 2\pi f_1 \tau_0 \tag{A.15}$$

$$= \phi_1 - \frac{f_1}{f_1 - f_2}(\phi_1 - \phi_2) \tag{A.16}$$

$$= \frac{f_2}{f_1 - f_2}\phi_1 + \frac{f_1}{f_1 - f_2}\phi_2. \tag{A.17}$$

Note that at that point, the phases of the two tones are the same and equal ϕ_0. Having the value of ϕ_0 we can find l by considering that:

$$2\pi f_1 l T_p = \phi_0 + 2n_0\pi, \tag{A.18}$$

under the condition that l and n_0 are the smallest possible integer numbers.

Once the value of l is found, the required delay can be calculated by using eq. (A.11):

$$\tau = \left(l + \frac{\phi_1 - \phi_2}{2\pi}\right) T_p \tag{A.19}$$

or

$$\tau = \left(l + \frac{\phi_1 - \phi_2}{2\pi}\right) (f_1 - f_2). \tag{A.20}$$

Bibliography

[1] G. Gibiino, A. Santarelli, S. Farsi, M. Myslinski, G. Avolio, and D. Schreurs, "S-functions mixer modeling for linearization purposes," in *European Microwave Integrated Circuits Conference (EuMIC)*, pp. 490–493, Oct 2011. pages xiv, 12, 29, 30, 31, 32, 34, 35, 36, 37, 152

[2] S. Farsi, J. Dooley, K. Finnerty, D. Schreurs, B. Nauwelaers, and R. Farrell, "Cross-memory polynomial modeling for RF circuits," in *7th European Microwave Integrated Circuits Conference (EuMIC)*, pp. 262–265, Oct 2012. pages xiv, xxiii, 29, 36, 37, 39, 42, 43, 123, 133, 152

[3] G. Cunha, S. Farsi, B. Nauwelaers, and D. Schreurs, "An FPGA-based digital predistorter for RF power amplifier linearization using cross-memory polynomial model," in *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits*, April 2014. pages xv, xxiii, 29, 44, 45, 47, 48, 49, 152

[4] S. Farsi, J. Dooley, K. Finnerty, D. Schreurs, B. Nauwelaers, and R. Farrell, "Behavioral modeling approach for array of amplifiers in active antenna array system," in *IEEE Topical Conference on Power Amplifiers for Wireless and Radio Applications (PAWR)*, pp. 73–75, Jan 2013. pages xv, 29, 48, 52, 54, 55, 152

[5] S. Farsi, P. Draxler, H. Gheidi, B. Nauwelaers, P. Asbeck, and D. Schreurs, "Characterization of intermodulation and memory effects using offset multisine excitation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, pp. 645–657, March 2014. pages xv, xvi, xvii, 57, 58, 60, 62, 66, 68, 70, 72, 73, 74, 76, 77, 83, 84, 120, 153

[6] S. Farsi, K. Remley, B. Nauwelaers, and D. Schreurs, "Efficient dithering technique with periodic waveforms for RF test and characterization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 3998–4007,

Nov 2013. pages xvii, xviii, 85, 91, 93, 94, 95, 96, 98, 99, 100, 101, 102, 103, 105, 107, 108, 109, 153

[7] S. Farsi, H. Gheidi, H.-T. Dabag, P. Gudem, D. Schreurs, and P. Asbeck, "Modeling of deterministic output emissions of power amplifiers into adjacent receive bands," *IEEE Transactions on Microwave Theory and Techniques*, *accepted*. pages xviii, xix, xx, 113, 116, 117, 119, 120, 122, 125, 127, 132, 136, 137, 138, 139, 140, 142, 154

[8] A. Tehrani, H. Cao, S. Afsardoost, T. Eriksson, M. Isaksson, and C. Fager, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 1510–1520, June 2010. pages xxiii, 22, 37, 38, 40, 50, 123

[9] A. Abidi, "The path to the software-defined radio receiver," *IEEE Journal of Solid-State Circuits*, vol. 42, pp. 954–966, May 2007. pages 2

[10] R. Bagheri, A. Mirzaei, M. Heidari, S. Chehrazi, M. Lee, M. Mikhemar, W. Tang, and A. Abidi, "Software-defined radio receiver: dream to reality," *IEEE Communications Magazine*, vol. 44, pp. 111–118, Aug 2006. pages 2

[11] S. Cripps, *RF power amplifiers for wireless communications, second edition*. The Artech House Microwave Library, Artech House, Incorporated, 2006. pages 4

[12] J. Pedro and N. Carvalho, *Intermodulation Distortion in Microwave and Wireless Circuits*. Artech House microwave library, Artech House, 2002. pages 4, 5, 20, 59, 61, 62

[13] D. Schreurs, M. O'Droma, A. Goacher, and M. Gadringer, *RF Power Amplifier Behavioral Modeling*. Cambridge University Press, 2009. pages 5, 20, 26, 61, 64

[14] W. Bosch and G. Gatti, "Measurement and simulation of memory effects in predistortion linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, pp. 1885–1890, Dec 1989. pages 8, 19

[15] O. Shimbo, "Effects of intermodulation, AM-PM conversion, and additive noise in multicarrier TWT systems," *Proceedings of the IEEE*, vol. 59, pp. 230–238, Feb 1971. pages 8, 19

[16] P. Roblin, D. Root, J. Verspecht, Y. Ko, and J.-P. Teyssier, "New trends for the nonlinear measurement and modeling of high-power RF transistors and amplifiers with memory effects," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, pp. 1964–1978, June 2012. pages 8

[17] J. Brinkhoff, *Bandwidth-dependent Intermodulation Distortion in FET Amplifiers*. Macquarie University (Division of Information & Communication Sciences, Department of Electronics), 2004. pages 8, 9

[18] J. Vuolevi, T. Rahkonen, and J. Manninen, "Measurement technique for characterizing memory effects in RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, pp. 1383–1389, Aug 2001. pages 9

[19] T. Cunha, E. Lima, and J. Pedro, "Validation and physical interpretation of the power-amplifier polar Volterra model," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 4012–4021, Dec 2010. pages 9, 23, 131

[20] S. Boumaiza and F. Ghannouchi, "Thermal memory effects modeling and compensation in RF power amplifiers and predistortion linearizers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 2427–2433, Dec 2003. pages 9

[21] A. Santarelli, F. Filicori, G. Vannini, and P. Rinaldi, "'backgating' model including self-heating for low-frequency dispersive effects in III-V FETs," *Electronics Letters*, vol. 34, pp. 1974–1976, Oct 1998. pages 9

[22] K. Kunihiro and Y. Ohno, "A large-signal equivalent circuit model for substrate-induced drain-lag phenomena in HJFETs," *IEEE Transactions on Electron Devices*, vol. 43, pp. 1336–1342, Sep 1996. pages 9

[23] J. Pedro, N. Carvalho, and P. Lavrador, "Modeling nonlinear behavior of band-pass memoryless and dynamic systems," in *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, pp. 2133–2136 vol.3, June 2003. pages 9

[24] B. Razavi, *RF Microelectronics*. Prentice Hall International Series in the Physical and Chemical Engineering Sciences, Pearson Education, 2011. pages 10

[25] P. Zhan, K. Qin, and S. Cai, "Joint compensation model for memory power amplifier and frequency-dependent nonlinear IQ impairments," *Electronics Letters*, vol. 47, pp. 1382–1384, December 2011. pages 10, 28

[26] H. Cao, A. Tehrani, C. Fager, T. Eriksson, and H. Zirath, "I/Q imbalance compensation using a nonlinear modeling approach," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, pp. 513–518, March 2009. pages 10, 28

[27] L. Anttila, P. Handel, and M. Valkama, "Joint mitigation of power amplifier and I/Q modulator impairments in broadband direct-conversion transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 730–739, April 2010. pages 10, 28

[28] C. Nader, P. Landin, W. Van Moer, N. Bjorsell, and P. Handel, "Performance evaluation of peak-to-average power ratio reduction and digital pre-distortion for OFDM based systems," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, pp. 3504–3511, Dec 2011. pages 10

[29] L. Guan and A. Zhu, "Low-cost FPGA implementation of Volterra series-based digital predistorter for RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 58, pp. 866–872, April 2010. pages 10

[30] W.-H. Yu, W.-F. Cheng, Y. Li, C.-F. Cheang, P.-I. Mak, and R. Martins, "Low-complexity, full-resolution, mirror- switching digital predistortion scheme for polar-modulated power amplifiers," *Electronics Letters*, vol. 48, pp. 1551–1553, November 2012. pages 10

[31] C. Yu, L. Guan, E. Zhu, and A. Zhu, "Band-limited Volterra series-based digital predistortion for wideband RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, pp. 4198–4208, Dec 2012. pages 10

[32] Y. Liu, W. Pan, S. Shao, and Y. Tang, "A new digital predistortion for wideband power amplifiers with constrained feedback bandwidth," *IEEE Microwave and Wireless Components Letters*, vol. 23, pp. 683–685, Dec 2013. pages 10

[33] Y. Ma, Y. Yamao, Y. Akaiwa, and K. Ishibashi, "Wideband digital predistortion using spectral extrapolation of band-limited feedback signal," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. PP, no. 99, pp. 1–10, 2014. pages 10

[34] Y.-J. Liu, W. Chen, J. Zhou, B.-H. Zhou, and F. Ghannouchi, "Digital predistortion for concurrent dual-band transmitters using 2-D modified memory polynomials," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 281–290, Jan 2013. pages 11

[35] J. Kim, P. Roblin, D. Chaillot, and Z. Xie, "A generalized architecture for the frequency- selective digital predistortion linearization technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 596–605, Jan 2013. pages 11

[36] M. Younes, A. Kwan, M. Rawat, and F. Ghannouchi, "Linearization of concurrent tri-band transmitters using 3-D phase-aligned pruned Volterra model," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 4569–4578, Dec 2013. pages 11

[37] B. Kim, I. Kim, and J. Moon, "Advanced Doherty architecture," *IEEE Microwave Magazine*, vol. 11, pp. 72–86, Aug 2010. pages 11

[38] J. Yan, C. Hsia, D. Kimball, and P. Asbeck, "Design of a 4-W envelope tracking power amplifier with more than one octave carrier bandwidth," *IEEE Journal of Solid-State Circuits*, vol. 47, pp. 2298–2308, Oct 2012. pages 11, 114

[39] N. Cheng and J. Young, "Challenges and requirements of multimode multiband power amplifiers for mobile applications," in *IEEE Compound Semiconductor Integrated Circuit Symposium (CSICS)*, pp. 1–4, Oct 2011. pages 11, 114

[40] P. L. Gilabert and G. Montoro, "Computationally efficient real-time digital predistortion architectures for envelope tracking power amplifiers," *International Journal of Microwave and Wireless Technologies*, vol. 5, no. 02, pp. 187–193, 2013. pages 11

[41] M. Hassan, L. Larson, V. Leung, D. Kimball, and P. Asbeck, "A wideband CMOS/GaAs HBT envelope tracking power amplifier for 4G LTE mobile terminal applications," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 60, pp. 1321–1330, May 2012. pages 11

[42] A. Kiayani, L. Anttila, and M. Valkama, "Digital suppression of power amplifier spurious emissions at receiver band in FDD transceivers," *IEEE Signal Processing Letters*, vol. 21, pp. 69–73, Jan 2014. pages 11, 123, 142

[43] M. Omer, R. Rimini, P. Heidmann, and J. Kenney, "A PA-noise cancellation technique for next generation highly integrated RF front-ends," in *IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*, pp. 471–474, June 2012. pages 11, 141, 142

[44] K. Eero, K. Pasi, H. Antti, H. Veli-Matti, P. Jani, T. Timo, S. Sandro, K. Tapio, and S. Hans, "Small size receiver band self-interference cancellation amplifier for 4g," in *42nd European Microwave Conference (EuMC)*. pages 11, 142

[45] M. Omer, R. Rimini, P. Heidmann, and J. Kenney, "A compensation scheme to allow full duplex operation in the presence of highly nonlinear microwave components for 4 G systems," in *IEEE MTT-S International Microwave Symposium Digest (MTT)*, pp. 1–4, June 2011. pages 11, 123, 141, 142

[46] A. Roussel, C. Nicholls, and J. Wight, "Frequency agile RF feedforward noise cancellation system," in *IEEE Radio and Wireless Symposium*, pp. 109–112, Jan 2008. pages 11, 142

[47] H.-T. Dabag, H. Gheidi, S. Farsi, P. Gudem, and P. Asbeck, "All-digital cancellation technique to mitigate receiver desensitization in uplink carrier aggregation in cellular handsets," *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, pp. 4754–4765, Dec 2013. pages 11

[48] P. Draxler and P. Asbeck, "Statistical evaluation of finite length digital modulation sequences," in *67th ARFTG Conference*, pp. 259–261, June 2006. pages 11, 64

[49] N. Carvalho, K. Remley, D. Schreurs, and K. Card, "Multisine signals for wireless system test and design [application notes]," *IEEE Microwave Magazine*, vol. 9, pp. 122–138, June 2008. pages 11, 57, 97

[50] L. Aladren, P. Garcia, J. De Mingo, P. Carro, and C. Sanchez-Perez, "Performance comparison of training sequences for power amplifier linearization systems," in *Wireless Communication Systems (ISWCS), 2011 8th International Symposium on*, pp. 6–10, Nov 2011. pages 11

[51] C. P. Silva, C. Clark, A. A. Moulthrop, and M. S. Muha, "Survey of characterization techniques for nonlinear communication components and systems," in *IEEE Aerospace Conference*, pp. 1713–1737, March 2005. pages 17

[52] J. Verspecht and D. Root, "Polyharmonic distortion modeling," *IEEE Microwave Magazine*, vol. 7, pp. 44–57, June 2006. pages 17, 18, 19, 31

[53] D. Root, J. Verspecht, D. Sharrit, J. Wood, and A. Cognata, "Broad-band poly-harmonic distortion (PHD) behavioral models from fast automated simulations and large-signal vectorial network measurements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 3656–3664, Nov 2005. pages 18, 33

[54] J. Verspecht, J. Horn, L. Betts, D. Gunyan, R. Pollard, C. Gillease, and D. Root, "Extension of X-parameters to include long-term dynamic memory effects," in *IEEE MTT-S International Microwave Symposium Digest*, pp. 741–744, June 2009. pages 19, 31

[55] J. Verspecht, J. Horn, and D. Root, "A simplified extension of X-parameters to describe memory effects for wideband modulated signals," in *75th ARFTG Microwave Measurements Conference (ARFTG)*, pp. 1–6, May 2010. pages 19, 31

[56] A. A. M. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Transactions on Communications*, vol. 29, pp. 1715–1720, November 1981. pages 20

[57] G. Zhou, H. Qian, L. Ding, and R. Raich, "On the baseband representation of a bandpass nonlinearity," *IEEE Transactions on Signal Processing*, vol. 53, pp. 2953–2957, Aug 2005. pages 21, 23, 124

[58] J. Demmel, *Applied Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, 1997. pages 22, 128, 129, 133

[59] D. Silveira, P. Gilabert, A. dos Santos, and M. Gadringer, "Analysis of variations of Volterra series models for RF power amplifiers," *IEEE Microwave and Wireless Components Letters*, vol. 23, pp. 442–444, Aug 2013. pages 22

[60] A. Zhu, J. Pedro, and T. R. Cunha, "Pruning the Volterra series for behavioral modeling of power amplifiers using physical knowledge," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, pp. 813–821, May 2007. pages 22

[61] J. Kim and K. Konstantinou, "Digital predistortion of wideband signals based on power amplifier model with memory," *Electronics Letters*, vol. 37, pp. 1417–1418, Nov 2001. pages 22, 24, 37, 50, 51, 79, 123, 124

[62] O. Hammi, F. Ghannouchi, and B. Vassilakis, "A compact envelope-memory polynomial for RF transmitters modeling with application to baseband and RF-digital predistortion," *IEEE Microwave and Wireless Components Letters*, vol. 18, pp. 359–361, May 2008. pages 22, 24, 133

[63] A. Zhu, P. Draxler, J. Yan, T. Brazil, D. Kimball, and P. Asbeck, "Open-loop digital predistorter for RF power amplifiers using dynamic deviation reduction-based Volterra series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, pp. 1524–1534, July 2008. pages 22, 25, 123, 133

[64] D. Morgan, Z. Ma, J. Kim, M. Zierdt, and J. Pastalan, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Transactions on Signal Processing*, vol. 54, pp. 3852–3860, Oct 2006. pages 22, 24, 25, 37, 50

[65] L. Ding and G. Zhou, "Effects of even-order nonlinear terms on power amplifier modeling and predistortion linearization," *IEEE Transactions on Vehicular Technology*, vol. 53, pp. 156–162, Jan 2004. pages 23, 129

[66] E. Lima, T. Cunha, H. Teixeira, M. Pirola, and J. Pedro, "Base-band derived Volterra series for power amplifier modeling," in *IEEE MTT-S*

International Microwave Symposium Digest, pp. 1361–1364, June 2009. pages 23, 129

[67] A. Kheirkhahi, J. Yan, P. Asbeck, and L. Larson, "Improved envelope injection and termination (EIT) RF power amplifier with envelope equalization for mobile terminal applications," *IEEE Transactions on Microwave Theory and Techniques*, vol. 62, pp. 341–351, Feb 2014. pages 27

[68] Y. Ma, Y. Yamao, Y. Akaiwa, and K. Ishibashi, "Wideband digital predistortion using spectral extrapolation of band-limited feedback signal," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. PP, no. 99, pp. 1–10, 2014. pages 27

[69] K.-F. Liang, J.-H. Chen, and Y.-J. Chen, "A quadratic-interpolated lut-based digital predistortion technique for cellular power amplifiers," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. PP, no. 99, pp. 1–5, 2014. pages 27

[70] M. Rawat, K. Rawat, F. Ghannouchi, S. Bhattacharjee, and H. Leung, "Generalized rational functions for reduced-complexity behavioral modeling and digital predistortion of broadband wireless transmitters," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, pp. 485–498, Feb 2014. pages 27

[71] S. N. Ba, *Efficient digital baseband predistortion for modern wireless handsets*. Georgia Institute of Technology, 2009. pages 27

[72] C. Xie, T. Zhang, and D. Liu, "Using X-parameters to model mixers," in *Microwave and Millimeter Wave Technology (ICMMT), 2012 International Conference on*, vol. 3, pp. 1–3, May 2012. pages 30

[73] J. Verspecht, D. Gunyan, J. Horn, J. Xu, A. Cognata, and D. Root, "Multi-tone, multi-port, and dynamic memory enhancements to PHD nonlinear behavioral models from large-signal measurements and simulations," in *IEEE MTT-S International Microwave Symposium Digest*, pp. 969–972, June 2007. pages 30

[74] E. Ifeachor and B. Jervis, *Digital Signal Processing: A Practical Approach*. Electronic systems engineering series, Prentice Hall, 2002. pages 38

[75] C. Eun and E. Powers, "A new Volterra predistorter based on the indirect learning architecture," *IEEE Transactions on Signal Processing*, vol. 45, pp. 223–227, Jan 1997. pages 44

[76] T. Cooper and R. Farrell, "Value-chain engineering of a tower-top cellular base station system," in *IEEE 65th Vehicular Technology Conference*, pp. 3184–3188, April 2007. pages 50

[77] J. McCormack, T. Cooper, and R. Farrell, "Tower-top antenna array calibration scheme for next generation networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, no. 2, pp. 11–11, 2007. pages 52

[78] J. Pedro and N. Carvalho, "Designing multisine excitations for nonlinear model testing," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 45–54, Jan 2005. pages 57

[79] K. Remley, "Multisine excitation for ACPR measurements," in *IEEE MTT-S International Microwave Symposium Digest*, vol. 3, pp. 2141–2144 vol.3, June 2003. pages 57, 106

[80] M. Myslinski, K. Remley, M. Mckinley, D. Schreurs, and B. Nauwelaers, "A measurement-based multisine design procedure," in *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits*, pp. 52–55, Jan 2006. pages 57

[81] K. Vanhoenacker, T. Dobrowiecki, and J. Schoukens, "Design of multisine excitations to characterize the nonlinear distortions during FRF-measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, pp. 1097–1102, Oct 2001. pages 57

[82] K. Vanhoenacker and J. Schoukens, "Detection of nonlinear distortions with multisine excitations in the case of nonideal behavior of the input signal," *IEEE Transactions on Instrumentation and Measurement*, vol. 52, pp. 748–753, June 2003. pages 58

[83] G. Vandersteen, G. Vandersteen, Y. Rolain, K. Vandermot, R. Pintelon, J. Schoukens, and W. Van Moer, "Quasi-analytical bit-error-rate analysis technique using best linear approximation modeling," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, pp. 475–482, Feb 2009. pages 58

[84] M. Weiss, C. Evans, and D. Rees, "Identification of nonlinear cascade systems using paired multisine signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 47, pp. 332–336, Feb 1998. pages 58

[85] D. Schreurs, M. Myslinski, and K. Remley, "RF behavioural modelling from multisine measurements: influence of excitation type," in *33rd European Microwave Conference*, vol. 3, pp. 1011–1014 Vol.3, Oct 2003. pages 64

[86] D. Schreurs and K. Remley, "Use of multisine signals for efficient behavioural modelling of RF circuits with short-memory effects," in *61st ARFTG Conference Digest*, pp. 65–72, June 2003. pages 64

[87] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951. pages 67

[88] J. Sevic, K. Burger, and M. Steer, "A novel envelope-termination load-pull method for ACPR optimization of RF/microwave power amplifiers," in *IEEE MTT-S International Microwave Symposium Digest*, vol. 2, pp. 723–726 vol.2, June 1998. pages 73

[89] J. Martins, N. Carvalho, and J. Pedro, "Intermodulation distortion of third-order nonlinear systems with memory under multisine excitations," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, pp. 1264–1271, June 2007. pages 73, 74, 79

[90] N. Kim, V. Aparin, and L. Larson, "Analysis of IM3 asymmetry in MOSFET small-signal amplifiers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, pp. 668–676, April 2011. pages 73, 74

[91] N. de Carvalho and J. Pedro, "A comprehensive explanation of distortion sideband asymmetries," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, pp. 2090–2101, Sep 2002. pages 74, 78

[92] P. Asbeck, H. Kobayashi, M. Iwamoto, G. Hanington, S. Nam, and L. Larson, "Augmented behavioral characterization for modeling the nonlinear response of power amplifiers," in *IEEE MTT-S International Microwave Symposium Digest*, vol. 1, pp. 135–138 vol.1, June 2002. pages 78

[93] J. Brinkhoff and A. Parker, "Effect of baseband impedance on FET intermodulation," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 1045–1051, Mar 2003. pages 78

[94] J. Brinkhoff, A. Parker, and M. Leung, "Baseband impedance and linearization of FET circuits," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 2523–2530, Dec 2003. pages 78

[95] C. Olsen, S. Sweeney, and K. Remley, "Representing bias-network baseband characteristics when simulating intermodulation distortion," in *76th ARFTG Microwave Measurement Symposium (ARFTG)*, pp. 1–6, Nov 2010. pages 78

[96] A. Zhu, J. Pedro, and T. Brazil, "Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 4323–4332, Dec 2006. pages 79

[97] A. Zhu, P. Draxler, C. Hsia, T. Brazil, D. Kimball, and P. Asbeck, "Digital predistortion for envelope-tracking power amplifiers using decomposed piecewise Volterra series," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, pp. 2237–2247, Oct 2008. pages 79, 80

[98] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 539–550, Apr 1999. pages 86

[99] B. Boser, K.-P. Karmann, H. Martin, and B. Wooley, "Simulating and testing oversampled analog-to-digital converters," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 7, pp. 668–674, Jun 1988. pages 86

[100] R. Gray and J. Stockham, T.G., "Dithered quantizers," *IEEE Transactions on Information Theory*, vol. 39, pp. 805–812, May 1993. pages 86

[101] E. Balestrieri, P. Daponte, and S. Rapuano, "A state of the art on ADC error compensation methods," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, pp. 1388–1394, Aug 2005. pages 86

[102] P. Carbone and D. Petri, "Effect of additive dither on the resolution of ideal quantizers," *IEEE Transactions on Instrumentation and Measurement*, vol. 43, pp. 389–396, Jun 1994. pages 86

[103] R. Skartlien and L. Oyehaug, "Quantization error and resolution in ensemble averaged data with noise," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, pp. 1303–1312, June 2005. pages 86, 88

[104] H. Pan and A. Abidi, "Spectral spurs due to quantization in nyquist ADCs," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, pp. 1422–1439, Aug 2004. pages 88

[105] R. Wannamaker, S. Lipshitz, J. Vanderkooy, and J. Wright, "A theory of nonsubtractive dither," *IEEE Transactions on Signal Processing*, vol. 48, pp. 499–516, Feb 2000. pages 89

[106] J. Vanderkooy and S. P. Lipshitz, "Resolution below the least significant bit in digital systems with dither," *Journal of the Audio Engineering Society*, vol. 32, no. 3, pp. 106–113, 1984. pages 100, 109

[107] P. Cruz, N. Carvalho, and K. Remley, "Evaluation of nonlinear distortion in ADCs using multisines," in *IEEE MTT-S International Microwave Symposium Digest*, pp. 1433–1436, June 2008. pages 101

[108] H. Fetterman, D. Martin, and D. Rich, "CMOS pipelined ADC employing dither to improve linearity," in *Proceedings of the IEEE Custom Integrated Circuits*, pp. 109–112, 1999. pages 101

[109] C. Hurrell, C. Lyden, D. Laing, D. Hummerston, and M. Vickery, "An 18 b 12.5 MS/s ADC with 93 dB SNR," *IEEE Journal of Solid-State Circuits*, vol. 45, pp. 2647–2654, Dec 2010. pages 101

[110] P. Hale, C. Wang, D. Williams, K. Remley, and J. Wepman, "Compensation of random and systematic timing errors in sampling oscilloscopes," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, pp. 2146–2154, Dec 2006. pages 106

[111] K. Remley, P. Hale, D. Williams, and C.-M. Wang, "A precision millimeter-wave modulated-signal source," in *IEEE MTT-S International Microwave Symposium Digest (IMS)*, pp. 1–3, June 2013. pages 106

[112] J. Martins, P. Cabral, N. Carvalho, and J. Pedro, "A metric for the quantification of memory effects in power amplifiers," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 4432–4439, Dec 2006. pages 108

[113] Z. Zhou and G. La Rue, "A 12-bit nonlinear DAC for direct digital frequency synthesis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 55, pp. 2459–2468, Oct 2008. pages 108

[114] J. Pereira, A. Cruz Serra, and P. Silva Girao, "Dithered ADC systems in the presence of hysteresis errors," in *Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference*, vol. 3, pp. 1648–1652 vol.3, 1999. pages 110

[115] F. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microwave Magazine*, vol. 10, pp. 52–64, Dec 2009. pages 123

[116] J. Moon and B. Kim, "Enhanced hammerstein behavioral model for broadband wireless transmitters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, pp. 924–933, April 2011. pages 123

[117] G. Jacovitti and G. Scarano, "Discrete time techniques for time delay estimation," *IEEE Transactions on Signal Processing*, vol. 41, pp. 525–533, Feb 1993. pages 127

[118] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 2006. pages 129, 133

[119] T. O'Sullivan, R. York, B. Noren, and P. Asbeck, "Adaptive duplexer implemented using single-path and multipath feedforward techniques with BST phase shifters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, pp. 106–114, Jan 2005. pages 142

Biography

Saeed Farsi received the B.Sc. and M.Sc. degrees from the University of Tehran, Tehran, Iran, in 2001 and 2004, respectively, and is about to receive the Ph.D. degree at the KU Leuven, Belgium in July 2014.

From 2004 to 2006, he was with the University of Tehran, as a Research Assistant. From 2006 to 2009, he was with the Fara-Afrand Company, Tehran, Iran, as a Design Engineer involved with high-power active/passive circuits in TV/radio transmitters. From November 2012 to June 2013, he was a Visiting Scholar with the University of California, San Diego (UCSD), CA, USA, where he worked on receive-band noise cancellation in RF transceivers by using DSP techniques. His current field of research is system-level modeling and linearization of RF/microwave circuits and systems.

List of Publications

Articles in International Journals

[1] **S. Farsi**, H. Gheidi, H.-T. Dabag, P. Gudem, D. Schreurs, and P. Asbeck, "Modeling of Deterministic Output Emissions of Power Amplifiers into Adjacent Receive Bands," *IEEE Trans. Microwave Theory Tech.*, accepted.

[2] **S. Farsi**, P. Draxler, H. Gheidi, B. Nauwelaers, P. Asbeck, and D. Schreurs, "Characterization of Intermodulation and Memory Effects Using Offset Multisine Excitation," *IEEE Trans. Microwave Theory Tech.*, vol. 62, no. 3, pp. 645-657, March 2014.

[3] **S. Farsi**, K. Remley, B. Nauwelaers, and D. Schreurs, "Efficient Dithering Technique with Periodic Waveforms for RF Test and Characterization," *IEEE Trans. Microwave Theory Tech.*, vol. 61, no. 11, pp. 3998-4007, Nov. 2013.

[4] **S. Farsi**, H. Ali Akbarian, D. Schreurs, B. Nauwelaers, and G. Vandenbosch, "Mutual Coupling Reduction in Antenna Arrays by Using a Simple Microstrip U-Section," *IEEE Antennas and Wireless Propagation Letters*, vol. 11, pp. 1501-1503, Nov. 2012.

[5] H.-T. Dabag, H. Gheidi, **S. Farsi**, P. Gudem, and Peter M. Asbeck, "All-Digital Cancellation Technique to Mitigate Receiver Desensitization in Uplink Carrier Aggregation in Cellular Handsets," *IEEE Trans. Microwave Theory Tech.*, vol. 61, pp. 4754-4765, Dec. 2013.

[6] P. J. Soh, B. Van den Bergh, H. Xu, H. Aliakbarian, **S. Farsi**, P. Samal, G. Vandenbosch, D. Schreurs, and B. Nauwelaers, "A Smart Wearable Textile Array System for Biomedical Telemetry Applications," *IEEE Trans.*

Microwave Theory Tech., vol. 61, No. 5, pp. 2253-2261, May 2013.

[7] H. Ali Akbarian, P. J. Soh, **S. Farsi**, H. Xu, G. Vandenbosch, D. Schreurs, E. Van Lil, and B. Nauwelaers, "Implementation of Project-Based Telecommunication Engineering Design Course," *IEEE Trans. Education*, vol. 57, pp. 25-33, Feb. 2014.

Articles in International Conferences

[1] G.C.L. Cunha, **S. Farsi**, B. Nauwelaers, and D. Schreurs, "An FPGA-Based Digital Predistorter for RF Power Amplifier Linearization Using Cross-Memory Polynomial Model," *International Workshop on Nonlinear Microwave and Millimetre-wave Integrated Circuits (INMMIC)* , Leuven, Belgium, Apr. 2014.

[2] **S. Farsi**, P. Draxler, H. Gheidi, B. Nauwelaers, P. Asbeck, and D. Schreurs, "Offset Multisine Excitation for Enhanced Characterization of Intermodulation and Memory Effects," *Power Amplifier Symposium*, San Diego, CA, USA, Sep. 2013.

[3] **S. Farsi**, J. Dooley, K. Finnerty, D. Schreurs, B. Nauwelaers, and R. Farrell, "RF Power Amplifier Behavioural Model Strategy for Active Antenna Array," *IEEE Topical Conference on RF Power Amplifiers for Wireless and Radio Applications (PAWR)*, pp. 73-75, Jan. 2013.

[4] **S. Farsi**, J. Dooley, K. Finnerty, D. Schreurs, B. Nauwelaers, and R. Farrell, "Cross-Memory Polynomial Modeling for RF Circuits," *European Microwave Integrated Circuits Conference (EuMIC)*, pp. 262-265, Oct. 2012.

[5] **S. Farsi**, D. Schreurs, and B. Nauwelaers, "Octave–Bandwidth Band-Pass Filter with High Selectivity and Wide Highly-Suppressed Stopband," *International Conference on Microwaves, Radar and Wireless Communications (MIKON)*, pp. 771-774, May 2012.

[6] **S. Farsi**, D. Schreurs, and B. Nauwelaers, "A Low-Cost Linear Upconverter to Extend Frequency Range of Vector Signal Generator," *International Workshop on Nonlinear Microwave and Millimetre-wave Integrated Circuits (INMMIC)* , Vienna, Austria, pp. 1-4, Apr. 2011.

[7] G. P. Gibiino, A. Santarelli, **S. Farsi**, M. Myslinski, G. Avolio, and D. Schreurs "S-Functions Mixer Modeling for Linearization Purpose," *European*

Microwave Integrated Circuits Conference (EuMIC), Manchester, UK, pp. 490-493, Oct. 2011.

[8] **S. Farsi**, D. Schreurs, and B. Nauwelaers , "Octave–Bandwidth Band-Pass Filter with High Spurious Response Suppression," *URSI Forum*, Brussels, Belgium, May 2010.

[9] K. Mohammadpour-Aghdam, **S. Farsi**, H. Ali Akbarian, G. Vandenbosch, and R. Faraji-Dana, "Broadband Log-periodical Antenna with Omni-Directional Radiation Pattern in Horizontal Plane," 4^{th} *European Conference on Antennas and Propagation 2010. 4th European Conference on Antennas and Propagation (EurAAP)*, Barcelona, Spain, pp. 1-4, Apr. 2010.

[10] K. Mohammadpour-Aghdam, H. Ali Akbarian, **S. Farsi**, R. Faraji-Dana, G. Vandenbosch, and W. De Raedt, "Miniaturized Loaded Crossed Dipole Amtenna with Omni-directional Radiation Pattern in the Horizontal Plane," *European Microwave Conference*, Vol. 1-3, pp. 763-766, Rome, ITALY, Sep. 2009.

Presented Workshops

[1] **S. Farsi**, J. Dooley, D. Schreurs, and R. Farrell, "Volterra-series-based models for future high frequency communication systems," *European Microwave Week (EuMW) workshop* on "Recent Advancements in RF Power Amplifier Modeling and Predistortion," pp. 1-26, Amsterdam, The Netherlands, 29 Oct 2012.

[2] D. Schreurs, **S. Farsi**, and K. Remley, " Multi-tone as a character-ization tool towards RF linearisation" *IEEE International Wireless Symposium (IWS) Workshop* on "Digital Techniques for Power Amplifiers Linearity and Efficiency Enhancement," Beijing, China, 18 April 2013.

[3] K. Remley, **S. Farsi**, D. Schreurs, D. Williams, P. Hale, and J. Wang, "Baseband corrections for precision millimeterwave signal measurements," *IEEE International Microwave Symposium (IMS) Workshop* on "The Importance of Low-frequency Measurements on High-frequency Characterization," Seattle, WA, USA, 3 June 2013.

[4] **S. Farsi**, H. Gheidi, H.-T. Dabag, P. Gudem, P. Asbeck, and D. Schreurs, "Behavioral models for receive-band noise cancellation in RF transceivers,"

IEEE International Microwave Symposium (IMS) Workshop on "Efficient RF Design Using Practical Behavioral Models - Bridging the Gap Between Measurements and Simulations," Tampa Bay, FL, USA, 2 June 2014.

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
TELECOMMUNICATIONS AND MICROWAVES
Kasteelpark Arenberg 10 box 2444
B-3001 Heverlee
saeed.farsi@esat.kuleuven.be
http://www.esat.kuleuven.be